A Multi-Stage Optimization Model for Air Force Reserve Officer Training Corps Officer Candidate Selection

THESIS

Marisha T. Kinkle, Captain, USAF

AFIT-OR-MS-ENS-12-16

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

DISTRIBUTION STATEMENT A. APPROVED FOR PUBLIC RELEASE;

DISTRIBUTION IS UNLIMITED

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the United States Air Force, the United States Department of Defense or the United States Government. This is an academic work and should not be used to imply or infer actual mission capability or limitations.

AFIT-OR-MS-ENS-12-16

A MULTI-STAGE OPTIMIZATION MODEL FOR AIR FORCE RESERVE OFFICER TRAINING CORPS OFFICER CANDIDATE SELECTION

THESIS

Presented to the Faculty

Department of Operational Sciences

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Operations Research

Marisha T. Kinkle, BA

Captain, USAF

March 2012

DISTRIBUTION STATEMENT A. APPROVED FOR PUBLIC RELEASE;

DISTRIBUTION IS UNLIMITED

AFIT-OR-MS-ENS-12-16

A MULTI-STAGE OPTIMIZATION MODEL FOR AIR FORCE RESERVE OFFICER TRAINING CORPS OFFICER CANDIDATE SELECTION

Marisha T. Kinkle, BA
Captain, USAF

Approved:

______________________________ ______________
Matthew J. Robbins, PhD (Chairman) Date

______________________________ ______________
Darryl K. Ahner, PhD (Member) Date

AFIT-OR-MS-ENS-12-16

Abstract

The Air Force Reserve Officer Training Corps (AFROTC) faces a declining budget and increased enrollment, creating the necessity for improving officer candidate selection thorough the various stages of its commissioning program. Three critical stages have a major impact on the type of officer AFROTC commission. This research proposes a multi-stage model to evaluate three stages: 1) the high school scholarship allocation process, 2) the in-college scholarship allocation process, and 3) commissioning. Each stage is examined individually so that collectively AFROTC decision makers are able to meet commissioning goals. Stage one involves allocating scholarships to high school candidates using the index policy heuristic. Stage two involves examining which candidates should be awarded an enrollment allocation while taking into account the probabilities of the candidate completing field training (FT) and going on to commission. A logistic regression is used to estimate the probabilities of FT completion and commissioning given a candidate's demographic information and college performance. Stage two is examined using dynamic programming with a knapsack formulation. Stage three involves selecting the most qualified cadets to commission into the USAF and is examined using a knapsack approach.

This research enables AFROTC to shape the workforce during the commissioning program with respect to specialty, diversity, and cost requirements. In addition, it provides the decision maker with an effective means to select candidates at each stage of the commissioning program. Analysis conducted for stage one indicates that use of the index policy heuristic provide AFROTC a means to achieve higher quality at equal expense. Analysis conducted for stages two and three allow AFROTC to assess

changes in total quality when considering different commissioning policies.

AFIT-OR-MS-ENS-12-16

To my parents who continue to support and guide me... To my son who who keeps me motivated and reminds me why I do what I do everyday... To my Air Force friends and mentors who took time to encourage me and push me to keep doing my best... especially Major General Alfred K. Flowers, Ret., Colonel Allen J. Jamerson, Major Kenneth M. Mercier, and Captain Carina R. Harrison.

Thank You!

Acknowledgements

I would like to take this opportunity to thank my thesis committee and my sponsor for their assistance in the completion of my thesis. First, I thank my advisor, Major Matthew J. Robbins for his encouragement, patience, and direction. Next, I would like to thank my reader, Dr. Darryl K. Ahner for his assistance and reviewing my thesis. Finally I would like to thank the men and women of the Air Force Reserve Officer Training Corps for their help throughout the process.

Marisha T. Kinkle

Table of Contents

Page

List of Figures

List of Tables

Table Page

A MULTI-STAGE OPTIMIZATION MODEL FOR AIR FORCE RESERVE OFFICER TRAINING CORPS OFFICER CANDIDATE SELECTION

I. Introduction

Every year, the United States Air Force (USAF) projects officer accessions to meet future USAF needs. The backbone of the USAF is its personnel. Without personnel, the USAF would fail to accomplish its mission, to defend the United States and protect its interests through aerospace power [11]. The personnel structure requires that the right number and quality of officers be assessed to satisfy future leadership requirements. The right mix of officers must be available to ensure a broad range of daily operations are maintained [13].

The Air Force Reserve Officer Training Corps (AFROTC), the Air Force Academy, and the Officer Training Corps (OTS) are the three commissioning programs in the USAF. AFROTC supplies over half of the total officer accessions annually and up to 70% of the officers accessed into the Air Force's technical Air Force Specialty Codes (AFSC) [9]. AFSCs are specific codes used to group positions based on similarity of functions and requirements for knowledge, education, training, experience, ability, and other common criteria [3]. Technical AFSCs require specific Science, Technology, Engineering, and Math (STEM) baccalaureate degrees. Besides AFROTC producing the most officers with STEM degrees, AFROTC also produces the largest number of foreign language and nursing majors [13].

The mission of AFROTC is to develop quality leaders for the USAF [11]. AFROTC recruits, educates, and commissions officer candidates from the 144 colleges and universities that host an AFROTC program based on Air Force (AF) requirements [11]. With AFROTC experiencing high enrollment numbers and retention rates, it is im-

perative that AFROTC select the most qualified students during each milestone in the program.

The three major milestones in AFROTC are the High School Scholarship Selection Process (HSSP), awarding of enrollment allocations (EA), and AFSC assignment. The purpose of the scholarship program is to support the mission of AFROTC and provide an incentive to attract and retain officer candidates of high quality whose leadership potential, personal and physical qualities, and academic objectives meet AF accession objectives. The EA process determines which cadets are eligible to attend field training and enter into the Professional Officer Course (POC). Once students complete the POC course, they are commissioned into the USAF, and awarded an AFSC by the Air Force Personnel Center (AFPC).

This research focuses on the development of a multistage optimization model that selects officer candidates for high school scholarships, enrollment allocations, and AFSC selection in order to maximize expected total quality. The quality score is measured by the candidate's Air Force Officer Qualifying Test (AFOQT) score. Currently, AFROTC uses historical information as a means of determining the number of scholarships and EA to give out and to whom. Historical attrition rates are calculated using personnel data collected over previous years, then used to forecast requirements. However, with complicating factors such as economic conditions, increased retention, and college and military costs, AFROTC requires a more flexible and reliable model [13].

AFPC is responsible for the assignment of AFSCs to commissioned officer candidates. AFPC allocates the different AFSCs based on the needs of the Air Force, academic major, and student preference. This thesis offers a method of allocating scholarships based on the Air Force needs, academic major, and an officer candidate's quality score.

Other branches of the military have approached similar problems in different ways. Raymond focuses on determining the number of reenlistments necessary to satisfy future force requirements in the United States Marine Corps by analyzing personnel numbers and applying transition rates over a period of time [22]. Ali et. al. examine the assignment of Navy enlisted personnel with the complicating factor of en route training [4]. They determine the optimal assignment of personnel through the use of the assignment problem with specially structured side constraints [4].

This thesis describes a multistage approach for allocating scholarship resources to officer candidates. Techniques depend on the stage of the problem. The quiz problem is used to determine a near optimal policy for allocating high school scholarships to officer candidates. Dynamic programming is utilized to determine an optimal policy for the allocation of EA slots to the most qualified cadets by considering budget, likelihood of completing FT and POC, and number of slots available. The assignment of AFSCs is examined using a knapsack problem formulation.

This thesis is organized into five chapters. This chapter provides an introduction to the material. Chapter two provides a literature review of the AFROTC commissioning process and in depth discussion of the knapsack problem, dynamic programming, quiz problem, and logistic regression. Chapter three discusses the methodology utilized to address the research problem while taking into consideration the decision makers' inputs. Chapter four discusses results and presents findings related to the research objectives. Chapter five provides concluding comments and ideas for future research.

II. Literature Review

This chapter provides provides a detailed description of the Air Force Reserve Officer Training Corps (AFROTC) commissioning program and reviews of the techniques used to formulate and analyze the problems of interests in this thesis. Section 2.1 details presents the background and the major milestones in the program to include commissioning, the high school scholarship selection process (HSSP), the enrollment allocation (EA) process, and the assignment of Air Force Specialty Codes (AFSC) to commissioned cadets. Sections 2.2, 2.3, 2.4, and 2.5 present the different methods utilized to analyze each of the milestones. These methods include application of the quiz problem, dynamic programming, logistic regression, and the knapsack problem.

2.1 Air Force Reserve Officer Training Corps

The mission of AFROTC is to develop quality leaders for the USAF. The AFROTC commissioning process begins with recruitment of high school students into the program. Ideally, students enter into the program during their freshman or sophomore year in college and enter into one of the Aerospace Studies (AS) courses. The program is divided into two categories: the General Military Course (GMC) and the Professional Officer Course (POC). The GMC course is composed of AS100s and AS200s while the POC course consists of AS300s and AS400s. In order to transition from the GMC to the POC course, officer candidates must compete and be awarded an EA. Once awarded an EA, the candidate must attend and complete Field Training (FT) in order to be to enter the POC and be qualified for commission. Upon completion of the POC course, students are commissioned into the USAF and assigned an AFSC by the Air Force Personnel Center (AFPC). Figure 1 is a representation of the AFROTC commissioning process.

In order to recruit and train the best qualified commissioning candidates, AFROTC

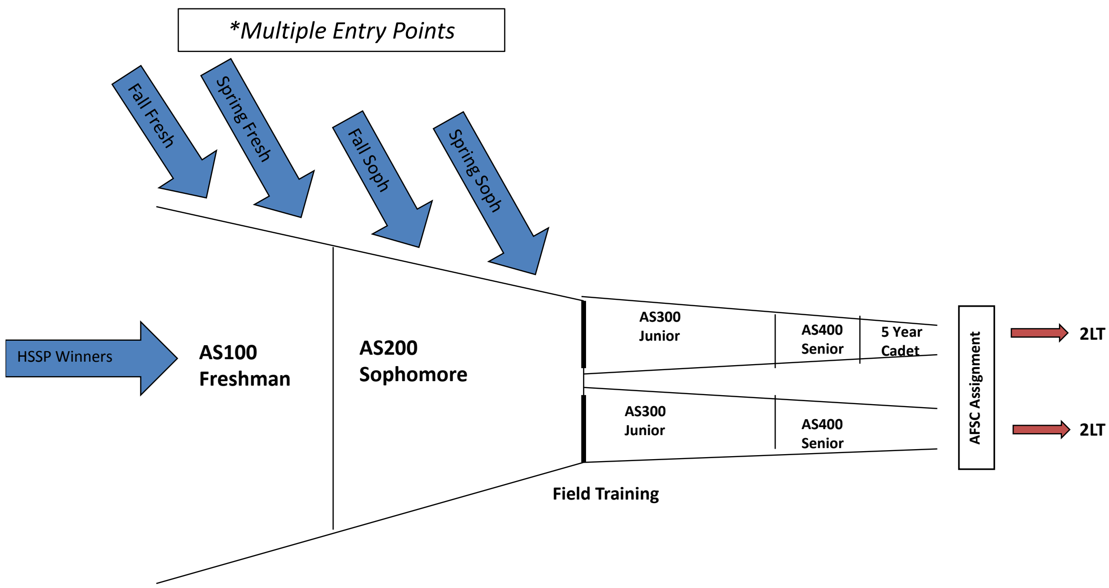


Figure 1. AFROTC Commissioning Process

uses scholarships as an incentive. The scholarship program consists of three main parts and has six distinct scholarship types. The program is authorized by Title 10, United States Code, Section 2107, Financial Assistance Program for Specially Selected Members, amendments to 10 U.S.C. 2107, and annual Nation Defense Authorization Acts [1]. AF/A1 provides requirements for officer production needs which guide the scholarship authorizations by academic specialty [1].

The AFROTC College Scholarship Program (CSP) contains the following three components: the High School Scholarship Program (HSSP), the In-College Scholarship Program (ICSP), and the Enlisted Commissioning Program (ECP) [1]. The

entire scholarship program is managed by the AFROTC Scholarships Branch (AFROTC/RRU). The component programs are managed by two offices within AFROTC/RRU, the High School Scholarships Section (AFROTC/RRUC) and the In-College and Enlisted Scholarships Section (AFROTC/RRUE).

The HSSP provides 3- and 4-year scholarship offers to high school seniors and graduates with no full-time college experience. If students have participated in a joint high school/college program prior to high school graduation, they are still eligible to apply for a scholarship. Typically, to be eligible for a scholarship consideration, an applicant must attain an un-weighted cumulative grade point average (CGPA) of 3.0 (as measured at the end of the junior year in high school) or higher and achieve either a Scholastic Assessment Test (SAT) total score of 1100 or an American College Test (ACT) composite score of 24.

AFROTC establishes the CSP application and selection process. The annual allocation of scholarships is determined based on fiscal considerations and AF production goals. From this information, the total amount, types, and academic categories are determined. The CSP awards 4-Year Type 1, 4-Year Type 2, and 4-Year Type 7 scholarships to select applicants. Students granted the 4-Year Type 7 scholarship have the option of converting to a 3-Year Type 2 scholarship [1].

The AFROTC CSP begins with the application period. The application period runs from 15 March to 1 December of each year [2]. Students are required to submit applications no later than 1 December using the on-line CSP application at www.afrotc.com. In order to become eligible to compete for a scholarship, each applicant must send in all required materials no later than 15 January [2]. The required information includes the on-line application, certified transcripts, physical fitness exam results and SAT/ACT scores [2]. The applicant must be at least 17 years old by the last day of the term in which the scholarship is activated and be a

United States citizen [2]. Once the application is received by AFROTC/RRUC, the information is put into a database to confirm an applicant's eligibility status [2]. If an applicant is determined eligible, he or she is scheduled for a personal interview with an Air Force officer [2].

Once the results of the interview are sent to and received by AFROTC/RRUC, the applicant's package meets the next available CSP board [2]. The board consists of a 3-member panel typically consisting of AFROTC detachment commanders and Air Liason Officers (ALOs) [2]. Official board results are normally released within four weeks of the conclusion of each board and RRUC notifies each applicant of the board result [2].

Historically, when evaluating applicants for a scholarship offer, the 3-member panel reviews each applicant's academic summary, interview results, resume, and extracurricular activity sheet [17]. Each board member then decides a maximum point value for each of the following areas for each applicant: Leadership, Motivation, Fitness, and Other (optional) [17]. Each area has a maximum point value of 34 points [17].

During the AS200 year, officer candidates compete for EAs. The number of EAs given out by AFROTC is based on AF/A1 fiscal year commissioning requirements while taking into consideration attrition rates. An award of an EA indicates a candidate is guaranteed commission as long as he.she successfully completes field training and the remaining two years of the program. Recently, the number of EAs have been cut due to cut backs in the USAF resulting in fewer cadets receiving EAs.

According to AFROTC Instruction 36-2011 [1], an eligible candidate is submitted for EA consideration by his/her detachment commander. Detachment commanders provide a unit commander ranking (UCR) and order of merit (OM) for each candidate. The UCR evaluates a candidates potential based on performance as a cadet in the program and based on a whole person concept. The OM is calculated by

weighting factors including the relative standing score (RSS), CGPA, physical fitness assessment (PFA) score, and SAT-R (highest score between AFOQT, SAT and ACT). Table 1 is a representation of the different components and weightings for the OM. Per AFROTC Instruction 36-2011, candidates who do not receive and EA are disenrolled from the program.

Per AFROTC instruction 36-2011 [1], Towards the end of the program, during a

Table 1. Order of Merit Factors

FACTORS	RANGE	MULTIPLIED BY	WEIGHT
RSS (Note 1)	5-10	5	50%
Cumulative GPA	2.0-4.0	5	20%
PFT	75-100	0.15	15%
SAT-R (Note 2)	650-1600	0.009375	15%
Notes: 1. The formula for calculation RSS is (10*((1-R/c)+0.5/C), where R = UCR and C = Class Size 2. The SAT-R is used only for selection processing			

candidate's AS400 year, he/she assigned an AFSC by AFPC via the AFROTC Form 53. AFSC classification is primarily based on the needs of the Air Force at the time of entry onto extended active duty (EAD). AFPC published a "target list" of higher need AFSCs for the subject fiscal year. Candidates use this "target list" to decided which AFSCs for which they may be qualified and for which they are interested in volunteering.

AFROTC is currently interested in developing techniques to better allocate scholarships and enrollment allocations. The motivation is to increase the quality of future officer candidates. Sister service ROTC programs are also interested in improving the allocation of scholarships. In 1999, RAND conducted a study of the Army ROTC scholarship program to offer alternate ways for its design [12]. This study examined the program from a financial and value perspective [12]. It examined different ways

Army ROTC could balance tuition costs and value of different academic institutions: private, prestigious, and other [12]. RAND's approach was first to examine lessons learned from the previous and tiered scholarship programs [12]. It drew on both quantitative and qualitative data [12]. With this data, they examined how different scholarship programs affected students attending varied valued academic institutions and how it impacted the quality of the officer produced [12]. The value of the academic institution was determined by evaluating the records of students that had graduated and commissioned from a certain type of institution [12]. The two measures considered were the officer's years of service and promotion rates [12].

From its analysis, RAND discovered that officers graduating from prestigious private schools are the only type that display significantly higher promotion rates at all grades in a standard 20-year career. However, sending students to these institutions is also the most expensive alternative [12]. Based on this information, RAND recommended four different types of scholarship programs that balance quality and tuition costs that differ based on the Army's priorities [12].

Civilian academic institutions are also concerned with ensuring that they enroll the most qualified students into their programs. Camarena-Anthony [8] examines scholarship allocation at Texas Tech University. Her research seeks to satisfy enrollment goals while achieving, simultaneously, other institutional objectives [8]. Likelihood of enrollment is predicted, initially, using a logisitic regression. This information is then used in a goal-programming model that seeks an optimal, merit-based scholarship allocation aligned with major institutional goals of academic quality and diversity [8]. The model provides decision makers with an effective way of distributing scholarships to incoming freshman.

2.2 The Quiz Problem

When considering how to allocate scholarships to the most qualified officer candidates, it is important to define an optimal policy for doing so. One way to define an optimal policy is through the use of the quiz problem, which is an example of a class of stochastic scheduling problems [6]. In its simplest form, the quiz problem involves an individual who is given a list of N questions to answer in any order desired. There is a probability p_i that the individual will get question i correct and receive reward v_i. The goal is to choose the optimal sequence of questions that results in the maximum expected reward.

The quiz problem can be thought of as a deterministic combinatorial problem, where one is seeking the goal of obtaining the optimal sequence in which to answer questions [6]. The simple form of the optimal solution to the quiz problem is deterministic; questions should be answered in decreasing order of $p_i v_i/(1 - p_i)$ [6]. When this policy is used in variations of the quiz problem where it is not necessarily optimal, it is referred to as an index policy [6]. The greedy policy answers questions in decreasing order of the expected reward $p_i v_i$ and is considered suboptimal because it does not consider the future loss associated with getting a question incorrect [6].

Before one can apply the quiz problem, it is important to understand why it yields an optimal policy. Let N denote the number of questions available, and M denote the maximum number of questions which may be attempted. Each question has an expected reward or value v_i and a probability of a correct answer, p_i [6]. There are time window or precedence constraints on the possible order of questions. The expected reward of a feasible question order $(i_1, \ldots, i_M)$ is $V(i_1, \ldots, i_M)$ where:

$$V(i_1, \ldots, i_M) = p_{i_1}(v_{i_1} + p_{i_2}(v_{i_2} + p_{i_3}(\ldots + p_{i_M} v_{i_M}) \ldots)). \tag{2.1}$$

If the question order $(i_1, \ldots, i_M)$ is infeasible, it is denoted:

$$V(i_1, ..., i_M) = -\infty. \tag{2.2}$$

When $M = N$, this is the classical quiz problem and all question orders are feasible. For this case, the optimal solution is obtained by using an interchange argument [6]. Let i and j be the kth and $(k+1)st$ questions on an optimally ordered list

$$L = (i_1, ..., i_{k-1}, i, j, i_{k+2}, \ldots, i_N). \tag{2.3}$$

Now, consider the list

$$L' = (i_1, \ldots, i_{k-1}, j, i, i_{k+2}, \ldots, i_N), \tag{2.4}$$

which is obtained from L by interchanging the order of questions i and j. When comparing the expected reward of L and L', the following result is [6]

$$\begin{aligned} E\{\text{reward of L}\} = & E\{\text{reward of}\{i_1, \ldots, i_{k-1}\}\} \\ & + (p_{i_1}, .., p_{i_{k-1}})(p_i v_i + p_i p_j v_j) \\ & + p_{i_1} ... p_{i_{k-1}} p_i p_j E\{\text{reward of}\{i_{k+2}, ..., i_N\}\}. \end{aligned} \tag{2.5}$$

Since L is optimally ordered, the following is obtained

$$E\{\text{reward of L}\} \geq E\{\text{reward of L'}\},$$

so it follows that

$$p_i v_i = p_i p_j v_j \geq p_j v_j + p_j p_i v_i \tag{2.6}$$

or equivalently

$$\frac{p_i v_i}{1 - p_i} \geq \frac{p_j v_j}{1 - p_j}. \tag{2.7}$$

From this, one can conclude in order to maximize the total reward, the question should be answered in decreasing order of $\frac{p_i v_i}{1-p_i}$; this yields the index policy [6].

During the high school scholarship and enrollment allocation selection processes, the decision maker's goal is to select the most qualified cadet for each opportunity. The HSSP process involves considering all applicants based on the different qualifiers mentioned previously and summarizing each cadet's achievement into one composite or quality score [17]. The composite scores help to determine each candidate's qualification. This relates to the quiz problem by assigning a probability of a scholarship offer being accepted (probability of getting a question correct) and quality of commissioned cadet (reward).

The EA selection process is similar to the HSSP process, however the quality score is determined differently. The Order of Merit (OM) or quality score for the EA selection process is a weighted multiple of the detachment commander's rating, cumulative GPA, Physical Fitness Test (PFT), and SAT equivalent score [21]. For this thesis, the analysis uses a student's Air Force Officer Qualifying Test (AFOQT) score due to data limitations regarding OM scores. The probability of completing field training and commissioning will be used to determine a near optimal policy for

selecting cadets for EAs.

2.3 Dynamic Programming Approach

Dynamic programming (DP) is a mathematical analysis technique where complex problems are broken down into simpler decisions that are solved in a sequence of steps or stages [16]. At each stage the outcome cannot be explicitly defined and may have some probability associated with a specific outcome. Usually, the goal is to minimize an undesirable cost. The goal is to balance some current cost with unknown future costs.

The basic DP model has two assumptions: (1) an underlying discrete-time dynamic system, and (2) a cost function that is additive over time [5]. The dynamic system considers the evolution of the decision variables over time. The state of the system changes from stage to stage as decisions are made. The system has the form [5]:

$$x_{k+1} = f_k(x_k, u_k, w_k), k = 0, 1, ..., N - 1, \tag{2.8}$$

where

k indexes discrete time,

x_k is the state of the system and summarizes past information that is relevant to future optimization,

u_k is the control or decision variable to be selected at time k,

w_k is a random parameter (also called the disturbance or noise depending on the context),

N is the horizon or number of times control is applied,

and f_k is a function that describes the system and in particular the mechanism by which the state is updated.

The total cost function is additive and is denoted:

$$J(x_N) + \sum_{k=0}^{N-1} g_k(x_k, u_k, w_k), \tag{2.9}$$

where g_k is the cost in stage k and $J(x_N)$ is a terminal cost incurred at the end of the process. Due to the presence of w_k, the cost is typically a random variable with some associated probability. Therefore the problem is formulated as an optimization of expected cost

$$E\left\{ g_N(x_N) + [\sum_{k=0}^{N-1} g_k(x_k, u_k, w_k)] \right\}, \tag{2.10}$$

where the expectation is with respect to the joint distribution of the random variables involved. The optimization is over the controls $u_0, u_1, ..., u_{N-1}$, where each control u_k is selected with some knowledge of the current state x_k [5].

2.4 Knapsack Problems

Once students are commissioned into the USAF, AFPC must decide how to allocate the students to the different AFSCs. Priority is given to students with technical majors in order to fill technical AFSCs first. A useful method for assigning the best qualified individuals to technical AFSCs is through the use of knapsack problems. The knapsack problem is a problem of combinatorial optimization where given a set N, consisting of n items j with profit p_j and weight w_j, and the capacity value c, the objective is to select a subset of N such that the total *profit* of the selected items is

maximized and the total weight does not exceed c [19].

Consider the following simple example of the knapsack problem from Keller et.al [19]. A hiker is packing his knapsack (or rucksack) for an intense hiking trail and must decide which items to take with him. He has a large number of items all of which have the potential to be very useful to him. Each item is assigned a number $j \in 1, \ldots, n$ and a certain profit, p_j, representing the benefit to the hiker. Each item also has a weight, w_j, which increases the load of his bag with each new item placed in the knapsack. The hiker would like to limit the total weight of his bag so he fixes the maximum load capacity to c.

A knapsack problem can be solved by obtaining a solution to the following linear integer programming formulation [19]:

$$\begin{aligned} \text{(KP) maximize } & \sum_{j=1}^{n} p_j x_j \\ \text{subject to } & \sum_{j=1}^{n} w_j x_j \leq c, \\ & x_j \in \{0,1\}, \text{j=1,...,n.} \end{aligned} \tag{2.11}$$

The optimal solution vector is denoted by $x^* = x_1^*$,...,x_n^* and the optimal solution value is denoted z^*. The set X^* denotes the optimal solution set (i.e. the set of items corresponding to the optimal solution vector) [19].

2.5 Logistic Regression

Regression analysis is a statistical technique that allows modeling of relationships between one or more independent indicator variables and response variables [20]. Logistic regression is used when the response is binary . [20] When examining the EA and HSSP process, the relationship between various indicator variables, such as

standardized test scores, grade point average, etc. and the response variable, officer candidate retention is analyzed. Logistic regression concepts are used to construct a predictive model.

Logistic regression differs from linear regression because it does not represent the response variable as a linear combination of the indicator variables. Logistic regression establishes a relation between the response and the predictors using the *logit* function as the dependent variable and modeling it as a linear function of the predictors [23].

The binary response y_i has a probability of success, π_i, given a certain certain independent variables χ_i. The probability can take on any value between 0 and 1 and is expressed by:

$$\pi_x = \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + \ldots + x_{ik}\beta_k. \tag{2.12}$$

This function does not guarantee that the probability, π_x, will fall between 0 and 1. Instead, the following form is used:

$$\pi_x = \frac{1}{1 + \exp^{\beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + \ldots + x_{ik}\beta_k}}. \tag{2.13}$$

A model with more than one predictor can be written as:

$$\pi_x = \frac{\exp^{\beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + \ldots + x_{ik}\beta_k}}{1 + \exp^{\beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + \ldots + x_{ik}\beta_k}}. \tag{2.14}$$

This thesis involves classifying cadets into a 0 or 1 response, where 1 indicates a student that completes field training and commissions into the USAF. When a student is classified in the 0 category, the student failed to complete FT and commission. The logistic regression model assigns new observations to one of the categories depending on what stage is being optimized.

The classification process involves a series of steps. First, the probability of belonging to a specific class is calculated then it is classified into a specific category based on a cutoff value [23]. Typically, the cutoff value is set a 0.5. When the probability is greater than 0.5, the case is classified as a 1 and 0 otherwise. The cutoff value can be adjusted to a different value depending on the event's probability.

The cutoff value is an indication of sensitivity and specificity that classifies a test result. The cutoff value is determined using the area under the receiver operating characteristic (ROC) curve. The area under the ROC curve, ranging from zero to one, indicates the model's ability to discriminate between the subjects who experience an outcome of interest. The optimal cutoff value is one that maximizes both sensitivity and specificity.

With logistic regression, there are many assumptions that do not hold when compared with linear regression. The errors or residuals in the model will not have a normal distribution [20]. Since the response variable takes on the value of 0 or 1, the distribution would approximate to the binomial distribution [23]. Also, the the assumption of constant variance is violated [20]. The variance is a function of the mean. This means a higher variance will occur when $pi_x \approx 0.5$.

This thesis implements the statistical packages MINITAB and JMP to perform the logistic regression analysis. Although other statistical packages, MINITAB and JMP were chosen due to ease of use and interpretation [8]. JMP is used to complete forward selection, backward elimination, and mixed stepwise logistic regression. MINITAB is used because it provides diverse goodness of fit statistics and diagnostic graphic capabilities that JMP does not [8].

MINITAB offers five tests for examining goodness of fit: Pearson Chi-square, Deviance, Hosmer-Lemeshow, and two Brown tests. Each test provides a different interpretation of how well the logistic regression fits.

The Pearson Chi-square and deviance tests are the two most popular goodness of fit measures [14]. The Pearson test detects only major departures from the logistic response function by dividing the cases into unique classes with different combinations of the predictor variables and the groups [8]. Within each class, the replicated cases are of the same combination [8]. The expected number of responses in each category is calculated according to the logistic regression function. This determines the Chi-square goodness of fit statistic. The deviance test is based on comparison of the likelihoods of the fitted model and the full model [14]. Both models require multiple or repeated observations at all combinations of factors [8]. Since the sample used in this thesis may not necessarily meet this requirement, these two goodness of fit methods may not be appropriate.

The Hosmer-Lemeshow test is based on the grouping of estimated probabilities obtained from the fitted logistic model [14]. This test can be applied to unreplicated data and data with few replications [8]. This test assigns estimated probabilities, where the logit values are similar, into groups of risk which ensures that there are a fair number of observations in each group [14]. In order to determine if the logistic function is an appropriate fit, the chi-square statistic is examined against the cut-off value (chi-square statistic $\leq \chi^2(1-\alpha, c-2)$, where c is the number of different combinations of predictor variables) [8]. If the value of the test is greater that α, the model fit is appropriate [8].

MINITAB also provides two Brown goodness of fit tests: the alternative and symmetric alternative Brown tests. Brown examines the goodness of fit of a logistic regression model using a score test statistic [14].The general alternative test statistic, asymptotically distributed Chi-Square with two degrees of freedom, is determined

using the following:

$$T = s'C^{-1}s \tag{2.15}$$

where C is the estimated covariance matrix fro s, and $s' = (s_1, s_2)$, a vector of score statistics defined as the partial derivatives of the log likelihood $\left(\frac{dl}{dm_1}, \frac{dl}{dm_2}\right)$ and estimated in MINITAB as follows [8]:

$$s_1 = S(y_i = P(x_i))(1 + \frac{\log[P(x_i)]}{1 - P(x_i)}), \tag{2.16}$$

$$s_2 = S(y_i = P(x_i))(1 + \frac{\log[1 - P(x_i)]}{P(x_i)}). \tag{2.17}$$

The one degree of freedom test for the symmetric alternative is [8]:

$$\frac{(s_1 + s_2)^2}{\text{Var }(s_1 + s_2)}. \tag{2.18}$$

This test proved to perform better than the general alternative when the true model is symmetric [7]. The Brown statistic p-value is examined to conclude the model fit. A p-value less than α, indicates a lack of fit, hence the null hypothesis that the logistic model fits the data is rejected [8].

III. Methodology

3.1 Overview

This chapter provides a detailed description of the methodology used to optimize the selection of officer candidates during various stages in the AFROTC program. In order to address this problem, it is considered as a multi-stage model with three stages: 1) the high school scholarship allocation process, 2) the enrollment allocation process, and 3) Air Force Specialty Code (AFSC) allocation. Each stage must be examined individually so collectively AFROTC decision makers are able to meet commissioning goals. Stage one involves allocating scholarships to high school candidates. Stage two involves examining which candidates should be admitted into the POC while taking into account the probability that a student meet requirements for commission. Stage three involves selecting the most qualified cadets to commission into the USAF. Figure 2 is a representation of the AFROTC multistage process.

3.2 Stage One

Stage one involves selecting the most qualified high school officer candidates to award scholarships. A full analysis of stage one is not presented due to time constraints. Instead, a cursory analysis is presented using heuristics developed through the use of the classical quiz problem. This allows us to rank cadets based on the historical probability of certain students being offered a scholarship and a student's probability of remaining in the program until commission.

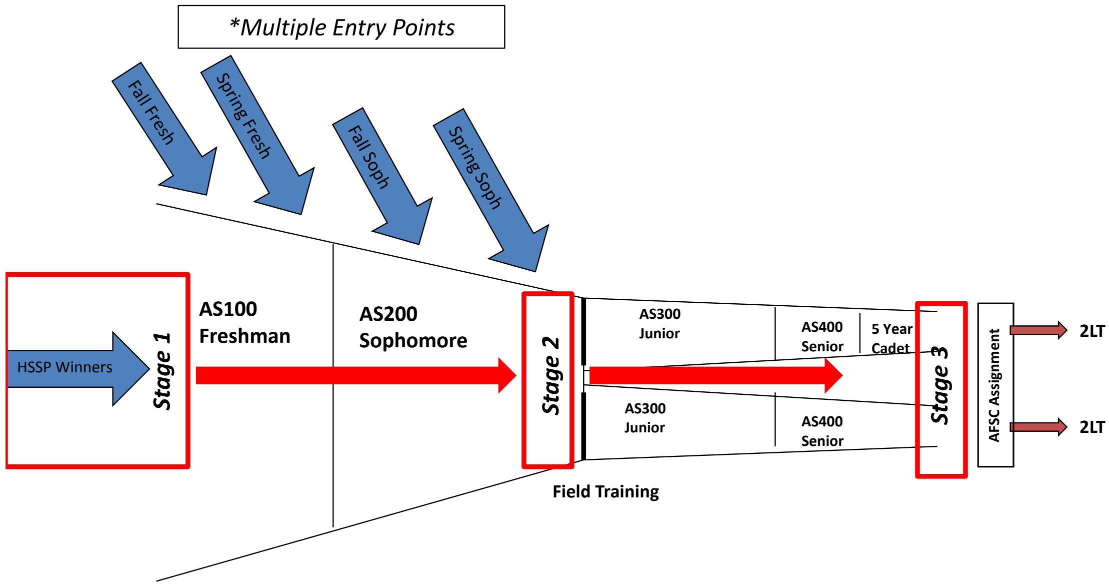


Figure 2. Multistage AFROTC Problem

3.3 Stage Two

Stage two focuses on the selection of officer candidates to receive an enrollment allocation (EA) which allows admission of the officer candidate into the POC. Each candidate's probability of successfully completing the program and becoming eligible for commission into the USAF must also be considered. This stage is examined using a d-dimensional knapsack problem with dynamic programming by evaluating students probability of completing field training and remaining in the program from the AS300 year through their final year in the ROTC program. Also, each cadet has a cost associated with continuing in the program. For this dynamic program, the state space is defined by the AFROTC budget, the number of POC applicants, and POC slots available.

Stage two is examined using LINGO software. LINGO is a comprehensive optimization software package developed by LINDO systems Inc. This software allows ease of model formulation, import/export of files into other programs, and powerful solvers.

The multi-dimensional knapsack problem can be viewed as a knapsack problem with a collection of different resource constraints or one constraint involving a multidimensional attribute [19]. Basically, the dimensionality of the knapsack problem refers to the number of constraints in the problem.

Decisions that consist of a series of interdependent stages resulting in a final decision are referred to as multiple-stage decisions problems [18]. These problems require the decision-maker to decide at each stage what action to take next in order to optimize performance at each stage [18]. Some examples include making decisions regarding working towards a degree, troubleshooting , medical treatment, scheduling, and budgeting [18].

Often, the method of backward induction or dynamic programming is used to solve such multi-stage problems [18]. Backward induction is the process of solving or examining a problem by working backwards in time to determine an optimal sequence of actions. For this multiple-stage decision problem, one begins by finding the optimal solution for the final stage and then proceeding backwards one stage at a time, finding the optimal solution at each stage, until the last (first) stage is complete. In order to utilize dynamic programming, all required assumptions must hold. In order to implement dynamic programming, one starts by solving a smaller sub problem of the d-KP and then extending the problem iteratively until the complete problem is solved [19]. After the various stages or periods are examined, the near optimal solution is a combination of near optimal solution to subproblems [19].

A near optimal solution to a knapsack problem, when an item or for this thesis, an officer candidate, is removed from the optimal knapsack packing, the remaining solution is near optimal for the subproblem defined by a decreased capacity and the new officer candidate set [19]. Making any other choice would decrease the optimal solution value.

The focus of stage two is maximizing the quality of the officer candidate subject to d-constraints. The base model examines the set of constraints related to the various AFSC requirements. The base model can be extended to examine other constraints concerning demographics such as an officer's candidate's region, sex, ethnicity, etc.

The model for stage 2 is a d-dimensional knapsack problem that seeks to maximize the overall quality of officer candidates. The model is represented as follows:

$$\text{Max} \qquad \sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij}P(C_i)Q_i \tag{3.1}$$

$$\text{subject to} \qquad \sum_{i=1}^{n} X_{ij} \leq \qquad N_j\text{,j=1,...,m} \tag{3.2}$$

$$X_{ij} \in \qquad 0,1\text{,i=1,...,n and j=1,...,m} \tag{3.3}$$

where decision variable $X_{ij} = 1$, if student i is awarded an EA slot during period j, 0 otherwise, $P(C_i)$ is the probability that a student will complete a certain portion of the program, Q_i is the quality score for student i, and N_j is the number of EA slots available during period j. This produces an optimal index policy $z_n(N_j)$. The objective, 3.1, seeks to maximize the total quality of the officers selected while taking into consideration a students probability that the student will satisfy the requirements to commission. Constraint 3.2 limits the number of of officer candidates awarded a EA slot. Constraint 3.3 indicates whether student i was awarded an EA slot during period j. There are n students and m periods

The probability that a student will satisfy commissioning requirements, $P(C_i)$ is

determined using a logistic regression on historical data. The probability is calculated then a student is assigned a 1 if it is likely the student will meet the requirement, 0 otherwise. For period/stage one, we assume students are assigned a one and N_0 is equal to the total number of officers required to meet AF/A1 goals. For period two, N_1, the probability of an AS300 going on to commission is considered and period three, N_2 focuses on whether or not an officer candidate will complete FT.

If $z_{j-1}(N_j)$ is known for all capacity values, $N_j = 0, \ldots, c$, then we can consider an additional item j and compute the corresponding solutions $z_j(N_j)$ by the following recursive formula

$$z_j(N_j) = \begin{cases} z_{j-1}(N_j) & \text{if } \; N_j < \sum_{i=1}^{n} \sum_{j=1}^{m} X_{ij} \\ \max \;\; z_{j-1}(N_j), z_{j-1}(N_j - \sum_{i=1}^{n} \sum_{j=1}^{m} X_{ij}) + \frac{P(C_i)Q_i}{1-P(C_i)} & \text{if } \; N_j \geq \sum_{i=1}^{n} \sum_{j=1}^{m} X_{ij} \end{cases} \tag{3.4}$$

The case $N_j < \sum_{i=1}^{n} X_{ij}$ means the considered knapsack is too small to contain item j at all. Therefore, item j does not change the optimal solution [19]. If item j does fit into the knapsack there are two possible choices. Either (1) the previous solution $z_{j-1}(N_j)$ remains unchanged or (2) adding item j to the knapsack improves the solution but decreases the capacity remaining. It is clear that the remaining capacity should be filled with the officer candidates that contribute the most quality.

3.4 Stage Three

The last decision point examined is the selection of the most qualified cadets to fill various AFSC requirements (Stage three). For this stage of the multistage problem, a d-dimensional knapsack problem (d-KP) is formulated and implemented in LINGO.

An integer program is used to find a solution to the d-KP, with the objective of maximizing the total quality. Once cadets are commissioned into the USAF, AFPC determines how to allocate the officer candidates into the varying career fields. AFPC makes these decisions based on requirements set forth by AF/A1. The basic model assumes that AFPC requires all AFSC requirements are met and the best qualified candidates are selected. Moreover, the model may be extended by considering other constraints or demographics such as an officer's regional background, sex, ethnicity, etc. When allocating these officer candidates into the varying career fields, these constraints determine the dimensions of the knapsack. For example, in the simple case, concerning only allocating scholarships to one career field, it is considered a one-dimensional knapsack problem (KP). When there are d-constraints, it is referred to as a d-dimensional knapsack problem (d-KP).

An integer program is a linear program in which at least one of the variables must take on an integer value [10]. When integer variables are restricted to 0 or 1 values, it is called a 0-1 (binary) integer program or binary IP [10]. A binary IP with a single $\leq$ constraint and positive objective function and constraint coefficients is referred to as a knapsack problem [10]. If the integer variables are not restricted to 0 or 1, it is referred to as an integer knapsack problem [10].

Stage three is formulated as follows:

$$\text{Max} \quad \sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij}Q_i \tag{3.5}$$

$$\text{subject to} \quad \sum_{i=1}^{n} X_{ij} \leq P_j, \quad \text{j=1,...,m} \tag{3.6}$$

$$\sum_{j=1}^{m} X_{ij} \leq 1 \tag{3.7}$$

$$X_{ij} \in \{0,1\}, \quad \text{i=1,...,n and j=1,...,m} \tag{3.8}$$

where the decision variable $X_{ij} = 1$, if student i is assigned to AFSC j, 0 otherwise, Q_i is the quality score for student i, and P_j is the required number of officers that must be assigned to AFSC j. The objective, 3.5, seeks to maximize the total quality of the officer candidates selected subject to constraints 3.6, 3.7, and 3.8. Constraint 3.6 limits the number of students assigned to each AFSC j. Constraint 3.7 ensures each candidate is only assigned one AFSC and constraint 3.8 is indicates whether student i was selected for AFSC j.

The quality of each candidate is indicated by his or her AFOQT score. A 2010 RAND study showed that the AFOQT is reasonable for predicting training success for a variety of Air Force officer specialties [15]. It is important to note that all officer candidates are not required to take the SAT or ACT and may not have a score in AFROTC's Web Intensive New Gain System (WINGS) database. Each officer candidate is required to take the AFOQT in order to be eligible for commission into the USAF and the score is maintained in WINGS.

IV. Implementation and Results

Chapter four presents the implementation and results for the multi-stage problem. Each stage requires a data mining process before the method is implemented. Using a dynamic programming approach, chapter four is presented working in decreasing order of the three stages. Analysis and results are presented for each stage.

4.1 Stage One

4.1.1 Data Mining Process.

AFROTC maintains a database of scholarship applicants in WINGS. During the application process candidate information is input into the WINGS database and stored for future use during the scholarship boards. AFROTC/RR provided an excel spreadsheet of all applicants who applied from 2001 through 2006 with the exception of 2003. The information for 2003 is unavailable due to a system malfunction that erased the information. This file contains an identification number of each applicant as well as 43 attributes including whether or not a student was offered and/or accepted a scholarship. This file is stored as 2001-2006_(HSSP_Application_and_Selection).xls in a file entitled Scholarship Applicants.

Next, the file is broken into five different files by application year: 2001 Applicants.xls, 2002 Applicant.xls, 2004 Applicants.xls, 2005 Applicants.xls, and 2006 Applicants.xls. Next, each file is prepared for analysis. Using the vlookup function in excel and the cadet commissioned data file, a new column was created to indicate whether a candidate went on to commission. The data is sorted using the eligible column to determine which students are eligible for scholarship. Only applicants eligible for scholarship are considered in this study.

Next, students are sorted from smallest to largest by scholarship offer. Applicants

may be boarded more than once during an application period. For this study, the analysis is based on the best scholarship offered. Next, the file is sorted by student identification number and duplicates are removed from the data set.

Finally student probabilities of scholarship acceptance and commission are calculate using two scoring metrics: a student's individual composite and SAT equivalent scores. The scores were separated into different groupings and the conditional probabilities are calculated based on his/her given score. These probabilities were calculated for three different files: FY 2003, FY2004, and FY 2001-2006. They are then combined and applied to FY 2004 data. These probabilities are found in the Appendix C.

4.1.2 Analysis.

Using the probabilities, the quiz problem formulation is implemented to determine the optimal policy for allocating scholarships. The probability of a student accepting a scholarship offer is equivalent to the probability of answering a question correctly and the probability of commission is equivalent to the reward. Each applicant's quiz score is calculated, then all scores sorted from largest to smallest. In order to make a comparison against AFROTC actual offers and commissionees, the quiz problem is utilized to determine the number of scholarship offers necessary to meet the same expected commissioning result. The average quality of commissioned applicants and total score are examined using the quiz policy. The estimated yearly scholarship cost are based on the average scholarship cost by type for FY2007. The costs are estimated using the actual offer and acceptance rate. This information can be found in Appendix C. Tables 2 and 3 indicate the results of the analysis.

As indicated in the literature review, AFROTC currently uses an applicant's in-

Table 2. Individual Composite Score Comparison

Data Used	# Apply	# Offers	#Accept	Scholarship Cost	Offer Avg Quality	# Comm	Comm Avg Quality	Total Quality
Actual	1620	293	94	$503,626	79	85	78.5	6,672.50
Historical (2001-2006 Averages)	1620	301	97	$517,342	88	85	87	7,395.00
Overlap		82			89	23	87	2,001.00
FY04 Data	1620	340	109	$584,374	75.64	85	73.78	6,271.30
Overlap		72			79.15	21	75.5	1,585.50
FY03 Data	1620	301	97	$517,342	88	85	85	7,225.00
Overlap		87			87	28	84	2,352.00

Table 3. SAT Comparison

Data Used	# Apply	# Offers	#Accept	Scholarship Cost	Offer Avg Quality	# Comm	Comm Avg Quality	Total Quality
Actual	1552	293	94	$503,626	1266	85	1246	105,910.00
Historical (2001-2006 Averages)	1552	292	94	$503,626	1405	85	1373	116,705.00
Overlap		85			1279	28	1251	35,028.00
FY04 Data	1552	360	115	$618,749	1275	85	1253	106,505.00
Overlap		85			1279	28	1251	35,028.00
FY03 Data	1552	292	94	$503,626	1404	85	1373	116,705.00
Overlap		69			1408	21	1377	28,917.00

dividual composite score to determine scholarship allocation. Using their existing method of assigning the individual composite score and combining it with the quiz problem, the results are compared with the actual offers. From Tables 2 and 3, it is seen that the average quality of student's who are offered a scholarship increases when the historical average and previous years probabilities are used. Using the historical average probabilities results in a 10.1% increase in the average quality of the applicant offered a scholarship. When examinging the index policy using the previous year's data the average quality of the applicant offered a scholarship increases by 7.6%. The average quality of the commissioned applicant increase by similar percentages with a 10.8% increase using historical data and 8.3% using the previous years data. Also, examining the estimated scholarship cost increase when the percentage of specific types of scholarships offered remains consistent with actual AFROTC offers. The historical (2001-2006 averages) result in the highest increase of the three categories. Similar results are indicated when using SAT scores as the quality score.

Stage one analysis using the index policy allows the decision maker to consider utilize a policy that not only takes into account an applicant's quality but to con-

sider his/her probability of accepting a scholarship offer and going on to commission. Given a student's quality score, individual composite or SAT score, the index policy is useful in allowing the decision maker to rank order applicants from highest to lowest and determine to whom to offer a scholarship. It also aids the decision maker in determining the number of scholarships to give out in order to meet a certain commissioning number.

4.2 Stage Two

4.2.1 Data Mining Process.

The data mining process for stage two is similar to the process done in stage three. The fall data pull files are used to compile all cadet information. For stage two, the data used is based on the selected indicator variables from the logistic regression which did the best job predicting whether an officer candidate goes on to commission or complete field training (FT). In order to complete the logistic regression, the data files are examined to determine which factors may contribute toward a student's completion of field training or commissioning.

4.2.1.1 Probability of Field Training Completion.

First, the probability of completing FT is examined. The ROTC Fall data pulls for FY 2006, 2008, 2009, and 2010, for all AS200 and AS250 students is extracted. FY 2007 data is left out for validation purposes. Once the information is extracted, it was compared with the FT selection file, extracted from the WINGS database, to determine which students are selected for FT using the vlookup function in excel. For each candidates's information it was compared with the next year's FY to see if the

student is enrolled in the program as an AS300. This information is used to indicate whether a student completed FT.

All officer candidates' information who were selected for field training is compiled into an excel spreadsheet named FY06toFY10_FT_woFY07.xls and saved in the "ROTC" data file. Each student has 74 different attributes describing their status in the program. Only factors that may contribute to an officer candidate's success in FT are kept: student id, region, AS level, Sex, Reserve Branch, Guard Branch, Active Duty Branch, CAP, Ethnicity, Race, Tech Major, CGPA, Scholarship Status, ACT, AFOQT, SAT, and AFPFT scores. Students who were missing CGPA,PFT, or had no score reported for either the SAT, ACT, or AFOQT are removed from the database. This resulted in 1,702 data points remaining.

From the remaining cadet records, the variables are analyzed to determine factors of interest. First, the variables are coded as categorical variables using Table 4. Next, the model is built to determine the candidate's probability of completing the program.

In order to build the model it is important to note the following considerations:

(1) The initial analysis includes all possible factors that may influence a candidate's completion of FT. The final model is built only on significant factors resulting from the logistic regression.

(2) The model is based on existing records from the AFROTC database for the selected FYs. For validation purposes, the model was tested using FY 2007 data.

(3) Only incomplete information was removed.

(4) In order to provide AFROTC with useful results, AFROTC personnel was included in the planning/information gathering process.

In order to fit the logistic regression, three selection methods are used for the variable selection process. The selection methods include forward selection, back-

Table 4. Variable Coding

Variable	Coding	Type	Count	Frequence(%)	Mean	Std Dev	Description
ID		Num					Student ID assigned to each candidate
Regn	NW = 1	Cat	1818	25.3			Numeric values assigned to each region
	NE = 2		1730	24.1			
	SW = 3		1687	23.5			
	SE = 4		1954	27.2			
AS Level	AS200 = 1	Cat	6842	95.2			Officer candidate's AS year
	AS250 = 2		347	4.8			
FT Comp	Completed = 1	DI	6457	89.8			Indicates completion of Field Training
	NonCompletion = 0		732	10.2			
Sex	Male = 1	Cat	5602	77.9			Assigns value to student's sex
	Female = 2		1587	22.1			
Race	American Indian = 1	Cat	44	0.6			Numeric values assigned to each category of racial group
	Asian = 2		476	6.6			
	Black = 3		406	5.6			
	Interracial = 4		140	1.9			
	Pacific Islander = 5		40	0.6			
	Unknown = 6		446	6.2			
	White = 7		5637	78.4			
Tech Major	Technical = 1	Cat	3303	45.9			Student has tech or nontech major
	Non-Technical = 2		3886	54.1			
Term GPA	N/A	Num			3.07	0.62	A student's term GPA at the beginning of the Fall when elgible for FT
Cumm GPA	N/A	Num			3.10	0.50	A student's cumlative GPA at beginning of Fall when elgibile to attend FT
On Schol	On Scholarship = 1	Cat	5450	75.8			Indicates whether a student is on scholarship during the Fall enrollment of FT eligibility year
	Non-Scholarship = 2		1739	24.2			
SAT-R					1195.54	173.03	Student's highest SAT equivalent score
AFPFT Score	N/A	Num			90.40	6.62	Student's most current physical fitness test score
Mil Experience	Yes = 1	Cat	288	4.0			Indicates whther student has any military experience or participated in CAP
	No = 2		6901	96.0			

ward elimination and mixed stepwise logistic regression. These procedures are based on the Wald statistic and its p-value and is examined using the software JMP. The forward selection process involves selection of predictors using univariate analysis. This involves testing each factor individually for a logistic fit in the first stage. The significant factors are used to construct a multivariate model. The backwards elimination approach builds an initial model with all possible factors and drops the least significant factor until only significant factors remain. Mixed stepwise logistic regression allows factors that were removed/added to be added/removed again until only significant factors remain.

All models are then compared within each stepwise regression, and the best are selected from each for comparison and goodness of fit. Then interaction variables are examined using the same techniques and the best model is chosen based on goodness of fit.

First, basic models are examined and the best models chosen. Next, the models

are examine using interaction variables. The basic models can be found in the Appendix B. Below are the results for forward selection (Table 5), backward elimination (Table 6), and mixed stepwise (Table 7) regression with interaction included. Models 4.4, 4.5, 5.7, and 6.4 are selected based their log-likelihood values. From this information, each model is built in MINITAB to obtain five goodness of fit test values: Pearson Chi-square, Deviance, Hosmer-Lemeshow, and two Brown tests, indicated in table 8. For this thesis only the Hosmer-Lemeshow and two Brown tests are used to determine goodness of fit. The Pearson Chi-square and Deviance tests are not used because they require multiple or repeated observations of the same values for all possible predictors. Since this cannot be guaranteed, these two goodness-of-fit tests may not be appropriate for this model.

Table 5. FT Forward Stepwise Logistic Regression w/ Interaction Results

Variable	Model 4.1	Model 4.2	Model 4.3	Model 4.4	Model 4.5
Tech Major	-	-	-	0.009	0.008
Cumulative GPA	0.000	0.000	0.000	0.000	0.000
Scholarship Status	-	0.000	0.000	0.000	0.000
AFPFT	0.000	0.000	0.000	0.000	0.000
*Tech Major * Tech Major*					
*Tech Major * CGPA*					
*Tech Major * Scholarship Status*				0.066	0.073
*Tech Major * AFPFT*					
*CGPA * CGPA*					
*CGPA * Scholarship Status*		0.012	0.012	0.010	0.010
*CGPA * AFPFT*	0.061	0.219	0.392	0.412	0.517
*Schoalrship Status * Scholarship Status*					
*Scholarship Status * AFPFT*					0.145
*AFPFT * AFPFT*			0.009	0.008	0.011
Log-Likelihood	-2247.814	-2224.708	-2220.864	-2216.174	-2215.116
Pearson Test (p-value)	0.097	0.342	0.289	0.451	0.409
Deviance Test (p-value)	1.000	1.000	1.000	1.000	1.000
Hosmer-Lemeshow (p-value)	0.201	0.306	0.642	0.513	0.901
Brown: general alt. (p-value)	0.001	0.001	0.074	0.061	0.056
Brown: symmetric alt. (p-value)	0.023	0.002	0.198	0.154	0.366

Examining the Hosmer-Lemeshow and two Brown tests, the first three models are greater than the acceptance criterion ($\alpha = 0.05$). This means each of the first three model's fit is appropriate and the logistic is the appropriate link function. Examining

Table 6. FT Backward Stepwise Logistic Regression w/ Interaction Results

Variable	Model 5.1	Model 5.2	Model 5.3	Model 5.4	Model 5.5	Model 5.6	Model 5.7
Tech Major	-	-	-	-	-	-	0.014
Cumulative GPA	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Scholarship Status	-	-	-	-	0.000	0.000	0.000
AFPFT	0.000	0.000	0.000	0.000	0.000	0.000	0.000
*Tech Major * Tech Major*	-	-	-	-	-	-	
*Tech Major * CGPA*	0.973						
*Tech Major * Scholarship Status*	0.063	0.059	0.057	0.061	0.072	0.064	
*Tech Major * AFPFT*	0.408	0.406	0.393	0.378			
*CGPA * CGPA*	0.525	0.525	0.478				
*CGPA * Scholarship Status*	0.011	0.011	0.009	0.007	0.007	0.005	0.007
*CGPA * AFPFT*	0.605	0.606					
*Schoalrship Status * Scholarship Status*	-	-	-	-			
*Scholarship Status * AFPFT*	0.144	0.144	0.127	0.117	0.123		
*AFPFT * AFPFT*	0.009	0.009	0.007	0.007	0.009	0.005	0.005
Log-Likelihood	-2214.547	-2214.548	-2214.681	-2214.936	-2215.325	-2216.507	-2218.214
Pearson Test (p-value)	0.312	0.317	0.361	0.434	0.472	0.525	0.474
Deviance Test (p-value)	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Hosmer-Lemeshow (p-value)	0.598	0.686	0.503	0.632	0.835	0.335	0.649
Brown: general alt. (p-value)	0.053	0.055	0.046	0.045	0.037	0.034	0.061
Brown: symmetric alt. (p-value)	0.834	0.832	0.537	0.287	0.203	0.068	0.126

Table 7. FT Mixed Stepwise Logistic Regression w/ Interaction Results

Variable	Model 6.1	Model 6.2	Model 6.3	Model 6.4	Model 6.5	Model 6.6
Tech Major	-	-	-			
Cumulative GPA	0.000	0.000	0.000	0.014	0.008	0.008
Scholarship Status	-	0.000	0.000	0.000	0.000	0.000
AFPFT	0.000	0.000	0.000	0.000	0.000	0.000
*Tech Major * Tech Major*					0.000	0.000
*Tech Major * CGPA*						
*Tech Major * Scholarship Status*					0.064	0.072
*Tech Major * AFPFT*						
*CGPA * CGPA*						
*CGPA * Scholarship Status*		0.012	0.012	0.007	0.005	0.007
*CGPA * AFPFT*	0.061	0.219	0.392			
*Schoalrship Status * Scholarship Status*						
*Scholarship Status * AFPFT*						0.123
*AFPFT * AFPFT*			0.009	0.006	0.005	0.009
Log-Likelihood	-2247.814	-2224.708	-2220.864	-2220.054	-2218.388	-2215.939
Pearson Test (p-value)	0.097	0.342	0.289	0.555	0.601	0.507
Deviance Test (p-value)	1.000	1.000	1.000	1.000	1.000	1.000
Hosmer-Lemeshow (p-value)	0.201	0.306	0.642	0.505	0.758	0.802
Brown: general alt. (p-value)	0.001	0.001	0.074	0.004	0.002	0.008
Brown: symmetric alt. (p-value)	0.023	0.002	0.198	0.011	0.005	0.086

Table 8. FT Completion Goodness of Fit Results

Variable	Model 4.4	Model 4.5	*Model 5.7*	Model 6.4
Tech Major	0.009	0.008	0.014	
Cumulative GPA	0.000	0.000	0.000	0.000
Scholarship Status	0.000	0.000	0.000	0.000
AFPFT	0.000	0.000	0.000	0.000
*Tech Major * Tech Major*				
*Tech Major * CGPA*				
*Tech Major * Scholarship Status*	0.066	0.073		
*Tech Major * AFPFT*				
*CGPA * CGPA*				
*CGPA * Scholarship Status*	0.010	0.010	0.007	0.007
*CGPA * AFPFT*	0.412	0.517		
*Scholarship Status * Scholarship Status*				
*Scholarship Status * AFPFT*		0.145		
*AFPFT * AFPFT*	0.008	0.011	0.005	0.006
Log-Likelihood	-2216.174	-2215.116	-2218.214	-2220.054
Pearson Test (p-value)	0.451	0.409	0.474	0.555
Deviance Test (p-value)	1.000	1.000	1.000	1.000
Hosmer-Lemeshow (p-value)	0.513	0.901	0.649	0.505
Brown: general alt. (p-value)	0.061	0.056	0.061	0.004
Brown: symmetric alt. (p-value)	0.154	0.366	0.126	0.011
Concordant Pairs(%)	68.8	68.8	68.7	68.6
Discordant Pairs (%)	30.4	30.4	30.5	30.5
Ties (%)	0.9	0.8	0.9	0.9

the two brown values of the last model, we see that it is not a fit. Of the first three models, it is seen that log-likelihood values vary minimally. The concordant pair's values also do not show much variation. Any of the first three models could be chosen for the predictive model. Since model 5.7 contains only significant variables, this model is chosen.

The estimated coefficients of the final model become the parameters in the logistic regression probability function. This allows the estimation of the probability of an officer candidate completing FT using the following:

$$\pi_x = 1 - \frac{\exp^{\beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + \ldots + x_{ik}\beta_k}}{1 + \exp^{\beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + \ldots + x_{ik}\beta_k}} \tag{4.1}$$

$$= 1 - \frac{\exp^{5.8901 + (0.2000)x_1 - (0.7897)x_2 - (0.6757)x_3 - (0.0589)x_4 - (0.4424)x_5 - (0.0008)x_6}}{1 + \exp^{5.8901 + (0.2000)x_1 - (0.7897)x_2 - (0.6757)x_3 - (0.0589)x_4 - (0.4424)x_5 - (0.0008)x_6}} \tag{4.2}$$

where x_1 = Tech Major 1, x_2 = CGPA, x_3 = Scholarship Status 1, x_4 = AFPFT Score, x_5 = CGPA*Scholarship Status, and x_6 = AFPFT Score*AFPFT Score.

Next, the model is validated using FY 2007 data. FY 2007 data was excluded from the observations used to build the model. Each officer candidate's predicted likelihood of completing FT is computed using the above equation. For the model a cutoff value of 0.69 determines whether a candidate completes FT. If the logistic response is greater than 0.69, the candidate is assigned a 1 indicating FT completion. If the logistic response is less than 0.69, the candidate is predicted not to complete FT and is assigned a 0. The cutoff value is selected using receiving operating characteristic (ROC) analysis in JMP.

The validation involves comparing these values with the actual values of FT completion. When the prediction and actual values match, there is an accurate prediction. When the values do not match, the prediction in incorrect. Table 9 provides a summary of the validation results. When FT completion is predicted, the prediction is correct approximately 89% of the time. However, when FT non-completion is predicted it is only correct 69% of the time. Overall, the model is able to predict over 95% of officer candidates that did complete FT. These results indicate this model may be useful for the prediction of FT completion.

Table 9. FT Completion Validation Results

Predicted	Actual	Frequency
1	1	1493
1	0	196
0	1	9
0	0	4

4.2.1.2 Probability of Commissioning.

Next, the probability of commissioning given a student is enrolled as an AS300 is examined. Using the ROTC Fall data pulls for FY 2006, 2008, and 2009, all AS300 data is extracted. FY 2007 data is left out for validation purposes. Once the information is extracted, it is compared with the commissioned file, extracted from the WINGS database, to determine which students commissioned using the vlookup function in excel. If a student is listed in the commissioned file, this indicates the student did commission.

All commissioned officer candidate's information is compiled into an excel spreadsheet named FY06toFY10_AS300_woFY07.xls and saved in the "ROTC" data file. Each student has 74 different attributes describing their status in the program. Only factors that may contribute to an officer candidate's success in FT were kept: student id, region, Sex, Reserve Branch, Guard Branch, Active Duty Branch, CAP, Ethnicity, Race, Tech Major, CGPA, Scholarship Status, ACT, AFOQT, SAT, and AFPFT scores. Students who are missing CGPA, PFT, or had no score reported for either the SAT, ACT, or AFOQT are removed from the database. This resulted in 6,357 data points remaining.

From the remaining cadet records, the variables are analyzed to determine factors of interest. First, the variables are coded as categorical variables using table 10. Next, the model is built to determine the candidate's probability of completing the program.

In order to build the model it is important to note the following considerations:

(1) The initial analysis includes all possible factors that may influence a candidate's probability of commissioning. The final model is built only on significant factors resulting from the logistic regression.

(2) The model is based on existing records from the AFROTC database for the

Table 10. Variable Categorical Coding

Variable	Coding	Type	Count	Frequence(%)	Mean	Std Dev	Description
ID		Num					Student ID assigned to each candidate
Regn	NW = 1	Cat	1818	25.3			Numeric values assigned to each region
	NE = 2		1730	24.1			
	SW = 3		1687	23.5			
	SE = 4		1954	27.2			
AS Level	AS200 = 1	Cat	6842	95.2			Officer candidate's AS year
	AS250 = 2		347	4.8			
FT Comp	Completed = 1	DI	6457	89.8			Indicates completion of Field Training
	NonCompletion = 0		732	10.2			
Sex	Male = 1	Cat	5602	77.9			Assigns value to student's sex
	Female = 2		1587	22.1			
Race	American Indian = 1	Cat	44	0.6			Numeric values assigned to each category of racial group
	Asian = 2		476	6.6			
	Black = 3		406	5.6			
	Interracial = 4		140	1.9			
	Pacific Islander = 5		40	0.6			
	Unknown = 6		446	6.2			
	White = 7		5637	78.4			
Tech Major	Technical = 1	Cat	3303	45.9			Student has tech or nontech major
	Non-Technical = 2		3886	54.1			
Term GPA	N/A	Num			3.07	0.62	A student's term GPA at the beginning of the Fall when elgible for FT
Cumm GPA	N/A	Num			3.10	0.50	A student's cumlative GPA at beginning of Fall when elgibile to attend FT
On Schol	On Scholarship = 1	Cat	5450	75.8			Indicates whether a student is on scholarship during the Fall enrollment of FT eligibility year
	Non-Scholarship = 2		1739	24.2			
SAT-R					1195.54	173.03	Student's highest SAT equivalent score
AFPFT Score	N/A	Num			90.40	6.62	Student's most current physical fitness test score
Mil Experience	Yes = 1	Cat	288	4.0			Indicates whther student has any military experience or participated in CAP
	No = 2		6901	96.0			

selected FYs. For validation purposes, the model was tested using FY 2007 data.

(3) Only officer candidates with incomplete information were removed.

(4) In order to provide AFROTC with useful results, AFROTC personnel was included in the planning/information gathering process.

In order to fit the logistic regression, similar to the FT completion probability analysis, three selection methods are used for the variable selection process. The selection methods include forward selection, backward elimination and mixed stepwise logistic regression. These procedures are based on the Wald statistic and its p-value and are examined using the software JMP.

Next, all models are compared within each stepwise regression, and the best are selected from each for comparison and goodness of fit. Then interaction variables are examined using the same techniques and the best model is chosen based on goodness of fit.

Below are the results for forward selection (Table 11), backward elimination (Ta-

ble 12), and mixed stepwise (Table 13) regression with interaction. Models 7.1, which also corresponds to models 8.7 and 9.4, and 10.2 are selected based on their log-likelihood values. From this information, each model is built in MINITAB to obtain five goodness of fit test values: Pearson Chi-square, Deviance, Hosmer-Lemeshow, and two Brown tests. For this thesis only the Hosmer-Lemeshow and two Brown tests are not to determine goodness of fit. These values are indicated in table 14. The Pearson Chi-square and Deviance tests are not be used because they require multiple or repeated observations of the same values for all possible predictors. Since this cannot be guaranteed, these two goodness-of-fit tests may not be appropriate for this model.

Table 11. Commissioned Forward Selection Stepwise Logistic Regression with Interaction Results

Variable	Model 10.1	Model 10.2
Region	0.000	0.000
Ethnicity		0.023
Mil Experience	0.007	0.007
*Region*Ethnicity*		
*Region*Mil Experience*	0.160	0.148
*Ethnicity*MilExperience*		
Log-Likelihood	-2922.031	-2919.185
Pearson Test (p-value)	0.242	0.410
Deviance Test (p-value)	0.253	0.238
Hosmer-Lemeshow (p-value)	0.926	0.963
Brown: general alt. (p-value)	0.242	0.929
Brown: symmetric alt. (p-value)	0.095	0.822

Examining the Hosmer-Lemeshow and two Brown test, both models are greater than the acceptance criterion ($\alpha = 0.05$). This means each model's fit is appropriate and the logistic is the appropriate link function. The log-likelihood values vary minimally. The concordant pair's values also do not show much variation. Either model could be chosen for the predictive model. Since model 7.1 only has significant variables, this model is chosen.

Table 12. Commissioned Backward Elimination Stepwise Logisitc Regression with Interaction Results

Variable	Model 11.1	Model 11.2	Model 11.3	Model 11.4
Region	0.000	0.000	0.000	0.000
Ethnicity	0.011	0.054	0.023	0.024
Mil Experience	0.014	0.014	0.007	0.004
*Region*Ethnicity*	0.635			
*Region*Mil Experience*	0.129	0.126		
*Ethnicity*MilExperience*	0.289	0.276	0.148	
Log-Likelihood	-2918.142	-2918.268	-2918.388	-2920.178
Pearson Test (p-value)	0.269	0.337	0.288	0.325
Deviance Test (p-value)	0.260	0.329	0.250	0.225
Hosmer-Lemeshow (p-value)	0.955	0.978	0.821	0.804
Brown: general alt. (p-value)	0.274	0.642	0.210	0.530
Brown: symmetric alt. (p-value)	0.427	0.706	0.521	0.451

Table 13. Commissioned Mixed Stepwise Logistic Regression with Interaction Results

Variable	Model 12.1
Region	0.000
Ethnicity	0.023
Mil Experience	0.007
*Region*Ethnicity*	
*Region*Mil Experience*	0.148
*Ethnicity*MilExperience*	
Log-Likelihood	-2919.185
Pearson Test (p-value)	0.410
Deviance Test (p-value)	0.276
Hosmer-Lemeshow (p-value)	0.963
Brown: general alt. (p-value)	0.929
Brown: symmetric alt. (p-value)	0.822

Table 14. **Commissioned Goodness of Fit Results**

Variable	*Model 7.1/8.7/9.4*	Model 10.2
Region	0.000	0.000
Sex		
Race		
Ethnicity	0.021	0.023
Tech Major		
Cumulative GPA	0.224	
Scholarship Status		
SAT-R		
AFPFT		
Mil Experience	0.004	0.004
*Region*Ethnicity*		
*Region*Mil Experience*		0.148
*Ethnicity*MilExperience*		
Log-Likelihood	-2919.409	-2919.185
Pearson Test (p-value)	0.026	0.410
Deviance Test (p-value)	0.273	0.238
Hosmer-Lemeshow (p-value)	0.146	0.963
Brown: general alt. (p-value)	0.113	0.929
Brown: symmetric alt. (p-value)	0.169	0.822
Concordant Pairs(%)	56.6	56.8
Discordant Pairs (%)	40.7	40.5
Ties (%)	2.8	2.7

The estimated coefficients of the final model become the parameters in the logistic regression probability function. This allows the estimation of the probability of an officer candidate completing FT using the following:

$$\pi_x = 1 - \frac{\exp^{\beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + \ldots + x_{ik}\beta_k}}{1 + \exp^{\beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + \ldots + x_{ik}\beta_k}} \tag{4.3}$$

$$= 1 - \frac{\exp^{-1.8349 + (0.2251)x_1 - (0.4319)x_2 - (0.3640)x_3 - (0.0854)x_4}}{1 + \exp^{-1.8349 + (0.2251)x_1 - (0.4319)x_2 - (0.3640)x_3 - (0.0854)x_4}} \tag{4.4}$$

where x_1 = Region, x_2 = Military Experience, x_3 = Ethnicity, and x_4 = CGPA.

Next, the model is validated using FY 2007 data. FY 2007 data is excluded from the observations used to build the model. Each officer candidate's predicted likelihood of commissioning is computed using the above equation. This model uses a cutoff value of 0.91 to determine whether a candidate commissions. If the logistic response is greater than 0.91, the candidate is assigned a 1 indicating the officer candidate will go on to commission. If the logistic response is less than 0.91, the candidate is

predicted not to commission and is assigned a 0. The cutoff value is selected using receiving operating characteristic (ROC) analysis in JMP.

The validation involves comparing these values with the actual values of commissioning. When the prediction and actual values match, there is an accurate prediction. When the values do not match, the prediction in incorrect. Table 15 provides a summary of the validation results. When commissioning is predicted, the prediction is correct approximately 81% of the time. However, when non-commission is predicted it is only correct 33% of the time. Overll, the model was able to predict over 99% of officer candidates that did commission. These results indicate this model may be useful for the prediction of commissioning which is the focus.

Table 15. Commissioned Validation Results

Predicted	Actual	Frequency
1	1	1766
1	0	421
0	1	6
0	0	3

4.2.2 Analysis.

Stage two analysis is implemented in LINGO. The input file consists of student eligible for entrance into the POC during FY2007. The input file consists of each students identification number, his/her detachment, estimated tuition rate, region, sex, ethnicity, race, major, cumulative grade point average, scholarship status, AFOQT aptitude score, AFPFT score, and a column indicating whether or not a candidate has any military experience.

The tuition amounts for students enrolled during FY2007 were obtained from HQ AFROTC and Holm Center staff. All tuition rates are assumed to be type 7 schol-

arships. These scholarships have a cap of $9,000. Since students from a detachment can come from multiple colleges/universities, the maximum tuition rate paid for one student was used from that year.

From the probability calculations, new columns indicating each candidate's probability of each event are added to the spreadsheet. LINGO is used to implement the knapsack problem using dynamic programming and seeks to optimize overall officer candidate selection. It starts with determining which candidates to select at Period three. Period three assumes all students will commission. Period two takes into account a candidate's probability of commission. Finally, period one utilizes the calculated probability of field training completion. The LINGO formulation can be found in Appendix A.

First, the basic model is examined. Each officer candidate decision variable is a binary variable and is labeled 1 - if the candidate is selected and 0 - if the student is not selected. There are three constraints: 1) there is a requirement for each AFSC, 2) an officer candidate can only be selected for one AFSC, and 3) an officer candidate cannot be assigned to an AFSC in which he/she is not eligible. Table 16 displays the results from the basic model.

Next, an extended form of the model is implemented. In addition to the con-

Table 16. Stage 2 Basic Model Results

	Period 3	Percentages	Period 2	Percentages	Period 1	Percentages
Overall Quality Score	99,155		83,790		91,487	
Mean Quality Score	79		77		76	
Female	280	22.2%	294	22.2%	316	22.7%
Hispanic	43	3.4%	45	3.4%	51	3.7%
Black	34	2.7%	36	2.7%	45	3.2%
Asian or Pacific Islander	66	5.2%	65	4.9%	71	5.1%
American Indian	8	0.6%	8	0.6%	8	0.6%
MultiRacial	12	1.0%	12	0.9%	13	0.9%
Projected Tuition Cost	$5,167,053.00		$5,469,853.00		$5,581,150.00	

straints in the basic model, three additional constraints concerning diversity are in-

cluded. These constraints include minimum percentage requirements on the number of female, racial minority, and Hispanic candidates selected. This allows the decision maker to have the option of taking into consideration diversity among officer candidates to reflect the diversity of American society. For this model, we assume that the decision maker requires 20% of the candidates to be female and/or minority and 5% Hispanic. Table 17 reflects the results of the model when ran with the new constraints. It is important to note that these are minimums. Some officer candidates decide not to disclose their racial/ethnic background. The LINGO program selects candidates who are identified in WINGS by a specific racial or ethnic group.

When comparing the two tables, these new requirements cause changes in the

Table 17. Stage 2 Extended Model Results

	Period 3	Percentages	Period 2	Percentages	Period 1	Percentages
Overall Quality Score	97,462		82,522		90,100	
Mean Quality Score	77		76		74	
Female	282	22.4%	299	22.6%	315	22.7%
Hispanic	64	5.1%	67	5.1%	70	5.0%
Black	101	8.0%	106	8.0%	112	8.1%
Asian or Pacific Islander	76	6.0%	80	6.0%	84	6.0%
American Indian	13	1.0%	14	1.1%	14	1.0%
MultiRacial	11	0.9%	11	0.8%	11	0.8%
Projected Tuition Cost	$4,951,385.00		$5,186,461.00		$5,500,031.00	

diversity mix and overall quality score during the various periods of the program. When the extended model is compared against the basic model, there is a decrease in the overall quality score during each period and in the percentage of female and Asian/Pacific Islander candidates. The overall quality scores for periods one, two and three decrease by 1.5%, 1.5%, and 1.7% respectively. These are small percentage decreases for increases in the diversity percentage mix of Hispanic, African-American, and American Indian Applicants.

Stage two analysis allows the decision makers to make multiple considerations when determining enrollment allocations. It allows the decision maker to optimize

enrollment allocation subject to a defined quality measurement. The decision maker is able to create and implement his/her own quality measure to determine enrollment allocation and implement it into the program. For example,AFROTC personnel may choose to use the SAT-R score in place of the AFOQT aptitude score as the quality measurement. Stage two also allows the decision maker to change/add constraints and conduct sensitivity analysis. For this analysis, the diversity requirements were determined based on attempting to mirror the officer candidate pool with the United States population of undergraduate students. AFROTC may want to change the constraints to have diversity mirror the USAF eligible commission population instead of the overall population. Additional constraints for possible consideration are budget and establishing a minimum number of students that must be selected from the detachments.

4.3 Stage Three

4.3.1 Data Mining Process.

During the fall of every fiscal year, Air Force Reserve Officer Training Corps AFROTC) pulls data from the WINGS database. Specific attributes for every cadet are saved and stored by the Holm Center Commander's Action Group (Holm Center/CCX). In order to determine the near optimal officer Air Force Specialty Code (AFSC) assignment, the AS400 data is extracted and scrubbed for the necessary fields.

To obtain the input data, the information is extracted from the AFROTC's WINGS database which requires a secure login. The secure login is obtained from Headquarters AFROTC. The file CADET_POOL_COMMISSION.xls is extracted from the website and saved to the ROTC data file. This file contains information for every

officer candidate that has commissioned from the AFROTC program since FY99. Once the file is extracted, all social security numbers are removed yet officer identification (ID) number remain. In order to examine stage three, the information is sorted and only fiscal year (FY) 2010 information is utilized. FY2010 data is saved in the same file as FY2010Comm.xls. The commissioned file is used because of the assumption that AFPC is provided with a similar list of officer candidates eligible for commission.

The FY2010 commissioned file contains 82 different fields for each cadet. For stage three, the required fields taken from the file are: student id, sex, region, Race Total, Ethnicity, Major Degree, Category Select, Aptitude, Verbal, and Quantitative. The student id is a unique identification number assigned to each student that enrolls in the AFROTC program. Sex is male or female. AFROTC is broken up into four regions: Northwest, Northeast, Southwest and Southeast. Race is broken out into seven categories by an assigned value: 1 - American Indian, 2 - Asian, 4 - Black, 8 - Native Hawaiian/Other Pacific Islander, 16 - White and 32 - Unknown/Decline to Respond. Students may also indicate more than one race. When this occurs the values are summed. Any value that is not equal to one of the above indicates the student is multiracial. Ethnicity has three categories: 1 - Hispanic, 2 - Non-Hispanic, and 3 - Unknown/Decline to Respond. There are over 1200 majors and each major is indicated by a four letter/number combination in the Major Degree column. Category select indicates in which category a cadet belongs and is indicated in Table 18. The aptitude, verbal, and quantitative columns indicate the score a student received from his/her AFOQT score for each category.

Next, the regional, sex, race, and ethnicity columns are formatted into variables. The sex column is transformed into 1s and 0s where 1 indicates female and 0 indicates

Table 18. **Category Select**

Indicator	Category
A	ABM
D	Dental
H	Physician Assistant
J	Physical Therapy
L	Legal
N	Navigator/Combat Support Officer
O	Line Officer (Non-Tech Major)
OT	Line Officer (Tech Major)
P	Pilot
Q	Nurse
R	Premedical
T	Occupational Therapy
U	Pharmacy
V1	UAV
X	Revoked

male. Ethnicity is indicated as follows: 1 - Hispanic and 0 - Non-Hispanic/Unknown. Race is as follows: 1 - American Indian, 2 - Asian, 3 - Black, 4 - Multiracial, 5- Native Hawaiian/Other Pacific Islander, 6 - Unknown/Decline to Respond and 7 - White. Regional information is broken out by 1 - northwest, 2 - northeast, 3 - southwest, and 4 - southeast.

Depending on a student's major and whether or not a student has been selected for a rated position, he/she is eligible for specific AFSCs. Students selected for rated positions are identified in the category select column. Rated positions include pilot, navigator, ABM, and UAV. Also medical and law students are specifically identified. This section focuses on optimizing AFSC selection for non-medical and non-legal commissioned officer candidates. Medical ans law students are removed from the data set. Students who are identified with rated slots are assumed to take on that AFSC and are not eligible for other AFSCs.

Four columns are added to the data set: AFSC1, AFSC2, AFSC3, and AFSC4. These columns indicate student AFSC eligibility. Each AFSC is assigned a numerical value indicated in Table 19. AFSCs that do not require a specific major are coded by the number 28. All other AFSCs were given a specific numeric value and

saved in the "'ROTC"' data file under AFSC.xls. Each AFSC is then matched up with majors in the data table extracted from WINGS and is saved in the same file as Wings_Major_File.xls. The vlookup function in excel is used to match up each student's major with eligible AFSCs.

In order to determine the near optimal AFSC allocation policy, stage three seeks to maximize quality score. The quality score is measured by a student's AFOQT score. An additional column is added to the data set entitled Overall AFOQT. Overall AFOQT is the sum of the aptitude, verbal, and quantitative portions of the AFOQT.

Finally, in order to prepare the data for import into LINGO, name ranges are assigned for various categories. A named range is assigned to student ids (student/studentnum), eligible AFSCs (AFSC), ethnicity (ethnicity), race (race), sex (sex), and quality score (quality). The final doument containing all FY2010 candidates is named All2010.xls and is saved into the ROTC data file. The output file is named stage3results.xls. This file identifies the quality score for each candidate and whether or not the candidate was selected.

4.3.2 Analysis.

Once the data is cleaned, LINGO is used to implement the knapsack formulation developed for stage three. In order to determine the number of officer candidates required for each AFSC, AF/A1 provided FY10 AFROTC requirements. AF/A1 provided a data file named AFPC Metric2010.xls. Included in this file are AFSC requirements and AFROTC production numbers. For this analysis, the values are compared against each other and the minimum number is chosen. This information was used as a guide to determine requirements for the analysis.

The requirement information is incorporated into the LINGO code. Initially, a

Table 19. AFSC Coding

Career Field	AFSC	MajorRequirement	Number
Financial Mgmt	65F/W	12 hrs	1
Cost Analyst	65WX	18 hrs	2
Contracting	64P	24 hrs	3
Cyber Space Warfare	17D	24 hrs tech (probable)	4
Aeronautical Engr	62EXA	Aeronautical Engr	5
C-E (Arch)	32EXA	Architecture	6
Astronautical Engr	62EXB	Astronautical Engr	7
Behav Sci/Human Factors	61BX	Behavioral Psychology	8
Chemist/Biologist	61CX	Chemistry	9
C-E (Civil)	32EXC	Civil Engr	10
Computer Engr	62EXC	Computer Engr	11
OSI	71Sx	Criminology	12
Electrical Engr	62EXE	Electrical Engr	13
C-E (EE)	32EXE	Electrical Engr	14
Acquisition Mgmt	63A	Engr, Math, Mgt, or 24 hrs	15
C-E(Envir)	32EXJ	Environmental Engr	16
Project Engr	62EXG	General Engr	17
C-E (Gen)	32EXG	General Engr	18
Operations Research Analyst	61AX	Math / Ops Research	19
Mechanical Engr	*62EXH*	Mechanical Engr	20
C-E (Mech)	32EXF	Mechanical Engr	21
Physics/Nuclear Engineer	61DX	Physics	22
Pilot	11	PreSelected	23
Navigator	12	PreSelected	24
Air Battle Managmt			25
Remote Piloted Aircraft			26
Weather	15W	Meteorology	27
All Others			28
Band	35BX	None	
Aircraft Maint	21A	None	
Mun/Missile Maint	21M	None	
Logistics Readiness	21R	None	
Security Forces	31PX	None	
Public Affairs	35P	None	
Force Support	38F	None	
Combat Control	13DXA	None	
Special Tactics	13DXB	None	
Air Liasion Officer	13L	None	
Air Field Operations	13M	None	
Space/Missile	13S	None	
Intelligence	14N	None	

basic model is utilized that determines how to optimally fill AFSC slots solely based on quality score. Each officer candidate selection decision is a binary variable and is labeled 1 if the candidate is selected for an AFSC and 0 if the student is not selected. There are three constraints: 1) there is a requirement for each AFSC, 2) an officer candidate can only be selected for one AFSC, and 3) an officer candidate cannot be assigned to an AFSC in which he/she is not eligible. It is important to note that certain AFSCs require a specific academic major. For example, an operations research analyst, 61A, must have an undergraduate degree in mathematics or operations research. The results of the basic model are reflected in Table 20.

From the results in Table 20, the average individual quality score is 92.29. 14.39% of officer candidates assigned to an AFSC are female. Minority officer candidates make up 14.48% of the selected candidates and 4.41% of the selected are hispanic candidates. All AFSC requirements are met.

Next, an extended form of the model is implemented. In addition to the constraints in the basic model, three additional constraints concerning diversity are included. These constraints include minimum percentage requirements on the number of female, racial minority, and Hispanic candidates selected. This allows the decision maker to have the option of taking into consideration diversity among officer candidates to reflect the diversity of American society. For this extended model, we assume that the decision maker requires 20% of the candidates to be female and/or minority and 5% Hispanic. Table 21 reflects the results of the model when ran with the new constraints. It is important to note that these are minimums. Some officer candidates decide not to disclose their racial/ethnic background. The LINGO program selects candidates who are identified in WINGS by a specific racial or ethnic group.

Table 20. Stage 3 Basic Model Results

Overall Quality Score			109,640.30
Average Individual Quality Score			92.29
Diversity			
		Number	Percentage
Female Officers		171	14.39%
Minority Officers		172	14.48%
Hispanic Officers		56	4.71%
AFSC Allocation			
Career	**AFSC**	**Required**	**Results**
Financial Mgmt	65F/W	10	10
Cost Analyst	65WX	0	0
Contracting	64P	15	15
Cyber Space Warfare	17D	39	39
Aeronautical Engr	62EXA	13	13
C-E (Arch)	32EXA	2	2
Astronautical Engr	62EXB	2	2
Behav Sci/Human Factors	61BX	3	3
Chemist/Biologist	61CX	5	5
C-E (Civil)	32EXC	6	6
Computer Engr	62EXC	11	11
OSI	71Sx	5	5
Electrical Engr	62EXE	33	33
C-E (EE)	32EXE	2	2
Acquisition Mgmt	63A	53	53
C-E(Envir)	32EXJ	2	2
Project Engr	62EXG	30	30
C-E (Gen)	32EXG	10	10
Operations Research Analyst	61AX	14	14
Mechanical Engr	*62EXH*	8	8
C-E (Mech)	32EXF	4	4
Physics/Nuclear Engineer	61DX	7	7
Pilot	11	511	511
Navigator	12	120	120
Air Battle Managmt	13	75	75
Remote Piloted Aircraft	18	12	12
Weather	15W	1	1
All Others		195	195

Table 21. Stage 3 Extended Model Results

Overall Quality Score			109,279.40
Overall Quality Score			109,279.40
Average Individual Quality Score			91.99
Diversity			
		Number	Percentage
Female Officers		238	20.03%
Minority Officers		238	20.03%
Hispanic Officers		60	5.05%
AFSC Allocation			
Career	**AFSC**	**Required**	**Results**
Financial Mgmt	65F/W	10	10
Cost Analyst	65WX	0	0
Contracting	64P	15	15
Cyber Space Warfare	17D	39	39
Aeronautical Engr	62EXA	13	13
C-E (Arch)	32EXA	2	2
Astronautical Engr	62EXB	2	2
Behav Sci/Human Factors	61BX	3	3
Chemist/Biologist	61CX	5	5
C-E (Civil)	32EXC	6	6
Computer Engr	62EXC	11	11
OSI	71Sx	5	5
Electrical Engr	62EXE	33	33
C-E (EE)	32EXE	2	2
Acquisition Mgmt	63A	53	53
C-E(Envir)	32EXJ	2	2
Project Engr	62EXG	30	30
C-E (Gen)	32EXG	10	10
Operations Research Analyst	61AX	14	14
Mechanical Engr	*62EXH*	8	8
C-E (Mech)	32EXF	4	4
Physics/Nuclear Engineer	61DX	7	7
Pilot	11	511	511
Navigator	12	120	120
Air Battle Managmt		75	75
Remote Piloted Aircraft		12	12
Weather	15W	1	1
All Others		195	195

When comparing the Tables 20 and 21, it can be seen that the AFSC requirements are still met while the diversity mix and overall quality score changes. The overall quality score and average individual quality score change minimally with a 0.3% decrease in both however, there is an increase in percentage of every diversity element. The overall percentage of females and minorities went up 5% and the percentage of Hipsanics increased by 0.35%.

Stage three analysis allows the decision maker to optimally select candidates to fill AFSC requirements when the number of officer applicants exceed the number of AFSC slots available. This happens in cases where the projected number of requirements, which are predicted 4-5 years out, decrease due to Air Force budgetary or end-strength issues. It also allows decision makers to consider diversity constraints to match the overall United States population that is commission eligible.

V. Conclusions and Future Research

5.1 Conclusions

The multi-stage problem was developed to assist the decision maker in the officer candidate selection process through 3 stages of the program: 1) the high school scholarship allocation, 2) enrollment allocation, and 3) AFSC selection processes. Stage one offers an effective optimization tool for allocating scholarships to applicants. The stage two optimization tool allows the decision maker to optimally select officer candidates for EAs while taking into consideration AF/A1 and/or diversity requirements. Stage three focuses on the optimal AFSC allocation policy while allowing the decision maker to consider diversity constraints.

Stage one is cursory analysis of the high school scholarship selection process. Currently the AFROTC scholarships branch uses applicants' individual composite score to determine scholarship application and is considering using SAT equivalent scores instead. Through the use of the quiz policy, an alternative method of scholarship allocation is developed. The main advantage of this method is that it allows the decision maker to consider an applicant's probability of accepting a scholarship and commissioning given his or her individual composite or SAT equivalent score when awarding scholarships. Using this method combined with the historical probabilities results in an overall quality increase of selected applicants 10.1% with an increase in the commissioned applicants' quality score of 10.8%.

For stage two, the inclusion of the logistic regression analysis allows insights into contributing factors toward Field training completion and commissioning. Significant factors determined to influence FT completion were whether the candidate is a technical major, CGPA, scholarship status, and physical fitness score. An officer candidate's region, ethnicity, CGPA, and military experience were significant factors

affecting a student's probability of commission. The outcome of the logistic regression analysis results in probabilities of field training completion and commissioning respectively. The output is used as an input to the dynamic programming model for stage two.

The officer candidate selection, provided by the dynamic program for stage two, is sensitive to changes in AF/A1 and diversity requirements. The decision maker is able to create and implement his/her own quality measure to determine enrollment allocation and implement it into the program. For example,AFROTC personnel may choose to use the SAT-R score in place of the AFOQT aptitude score as the quality measurement. Stage two also allows the decision maker to change/add constraints and conduct sensitivity analysis. Additional constraints for possible consideration are budget and establishing a minimum number of students that must be selected from the detachments. The optimal selection policy relies on the decision maker's priorities and preferences.

Stage three uses a knapsack problem approach to determine the optimal AFSC allocation based on a student's AFOQT aptitude score. Stage three analysis allows the decision maker to optimally select candidates to fill AFSC requirements when the number of officer applicants exceed the number of AFSC slots available. This happens in cases where the projected number of requirements, which are predicted 4-5 years out, decrease due to Air Force budgetary or end-strength issues. It also allows decision makers to consider diversity constraints to match the overall United States population that is commission eligible.

The main advantages of stages two and three are similar. They are flexible to differing situations by changing the parameters and/or constraints in the models. They can be applied to future fiscal years.

Each stage of the multi-stage problem had at least one limitation. In stage one,

the assumption is made that all applicant's are evaluated then offered scholarships at the same time. During the application process, AFROTC holds multiple boards and offers scholarships at the conclusion of each board. The number of scholarships remaining depend on the number of applicants that accept offers from previous boards.

The major limitation of stage two is that it is based solely on quantitative rather than qualitative data. One major component that AFROTC currently uses when determining EA allocation is the detachment commander ranking which is qualitative; this is not considered in the analysis. In addition, due to limited information, the tuition rates are maximums for each detachment and not actual tuition rates for an officer candidate's college/university. Also other costs associated with a candidate continuing in the program are not considered such as book or monthly allowance stipends.

When an officer becomes eligible to commission, his/her information is sent to AFPC for AFSC allocation. One limitation of the stage three analysis is it does not take into account officer candidate preference or detachment commander's recommendation for AFSC assignment. It is solely based on quantitative data.

Although there are limitations to the multi-stage study, it provides the decision maker(s) with useful information throughout each stage. Stage one allows the decision maker the option of taking into consideration the probability of an applicant commissioning in addition to his/her quality score. Stages two and three allows for ease of sensitivity analysis of overall quality when considering EA/AFSC allocation and diversity requirements.

This research provides an application for determining officer candidate selection along various stages of the multi-stage program using various techniques. Stage one utilized a simple heuristic approach for optimal scholarship allocation. Stages two and three offer a dynamic knapsack formulation approach supported with software

and tools to assist the decision-making process. The multi-stage model was formulated to be easily applied by personnel at Headquarters AFROTC.

5.2 Future Research

Each stage of the multi-stage process can benefit from additional research. Stage one is simply a cursory analysis of the high school scholarship allocation process. Other factors that should be studied are an applicant's probability of reaching stage two, with analysis of contributing factors, and the number of scholarship's necessary to ensure AF/A1 goals are met. Also, it would be useful to expand the research to include the in-college scholarship program. Also, dynamic programming could be utilized to consider the multitple boards held during the scholarship allocation process. Stage two can be extended to a multi-objective problem to include the objective of minimizing the overall cost or budget when allocation EAs. All three stages could benefit from value focused thinking analysis to determine how exactly AFROTC should evaluate a candidate for selection at different stages. Sensitivity analysis of the value and evaluation criteria may give insights into what can improve the way the quality of an officer candidate is measured. This allows the decision maker to determine importance of an officer candidates college/university of attendance, GPA, SAT equivalent scores, etc.

Appendix A. LINGO Code

```
! Stage 3 Analysis;
model:
title Sample AFSC;
sets:
      student:quality, gender, race, ethnicity, studentnum;
      choice;
      variables(student,choice): planned, AFSC ;
      jobs: Required;

endsets

data:
      choice = 1, 2, 3, 4;          !2 possible AFSC choices;
      student = @OLE('All2010.xls', 'student');
      AFSC = @OLE('All2010.xls', 'AFSC');
      quality = @OLE('All2010.xls', 'Score');
      gender = @OLE('All2010.xls', 'gender');
      race = @OLE('All2010.xls', 'race');
      ethnicity = @OLE('All2010.xls', 'ethnicity');
      studentnum = @OLE('All2010.xls', 'studentnum');
      Required = 10 0 15 39 13 2 2 3 5 6 11 5 33 2 53 2 30 10 14 8 4 7 511
120 75 12 1 195;
      TotalRequired = 1188;
enddata


max= @sum(student(i):
      @sum(choice(j):
      planned(i,j)*quality(i)));

!Constraints;
@for(jobs(j):                               !Meet AFSC requirements;
@sum(variables(i,k)|AFSC(i,k) #EQ# j:
      planned(i,k)) < Required(j));

@sum(variables(i,k)|gender(i) #EQ# 1:             !At least 20% female;
              planned(i,k)) >
               TotalRequired*0.2;

@sum(variables(i,k)|race(i) #EQ# 1:         !At least 20% minorities;
               planned(i,k)) >
               TotalRequired*0.2;

@sum(variables(i,k)|ethnicity(i) #EQ# 1:  !At least 5% Hispanic;
               planned(i,k)) >
               TotalRequired*0.05;
@for(student(i):                            !No more than one job per student;
      @sum(choice(k):
            planned(i,k)) < 1);

@for(variables(i,k):                        !Don't assign an AFSC that isn't a
choice;
      planned(i,k) < AFSC(i,k));
```

```
FinancialMgmt = @sum(variables(i,k)|AFSC(i,k) #EQ# 1:
      planned(i,k));
CostAnalyst = @sum(variables(i,k)|AFSC(i,k) #EQ# 2:
      planned(i,k));
Contracting = @sum(variables(i,k)|AFSC(i,k) #EQ# 3:
      planned(i,k));
CyberSpaceWarfare = @sum(variables(i,k)|AFSC(i,k) #EQ# 4:
      planned(i,k));
AeroEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 5:
      planned(i,k));
CEArch = @sum(variables(i,k)|AFSC(i,k) #EQ# 6:
      planned(i,k));
AstoEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 7:
      planned(i,k));
BehavSciHumanFac = @sum(variables(i,k)|AFSC(i,k) #EQ# 8:
      planned(i,k));
ChemistBiologist = @sum(variables(i,k)|AFSC(i,k) #EQ# 9:
      planned(i,k));
CECivil = @sum(variables(i,k)|AFSC(i,k) #EQ# 10:
      planned(i,k));
CompEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 11:
      planned(i,k));
OSI = @sum(variables(i,k)|AFSC(i,k) #EQ# 12:
      planned(i,k));
ElecEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 13:
      planned(i,k));
CEElecEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 14:
      planned(i,k));
AcqMngmt = @sum(variables(i,k)|AFSC(i,k) #EQ# 15:
      planned(i,k));
CEEnvEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 16:
      planned(i,k));
ProjectEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 17:
      planned(i,k));
CEGenEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 18:
      planned(i,k));
OpsAnalyst = @sum(variables(i,k)|AFSC(i,k) #EQ# 19:
      planned(i,k));
MechEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 20:
      planned(i,k));
CEMechEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 21:
      planned(i,k));
PhysicsNucEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 22:
      planned(i,k));
Pilot = @sum(variables(i,k)|AFSC(i,k) #EQ# 23:
      planned(i,k));
Navigator = @sum(variables(i,k)|AFSC(i,k) #EQ# 24:
      planned(i,k));
ABM = @sum(variables(i,k)|AFSC(i,k) #EQ# 25:
      planned(i,k));
RemotePilot = @sum(variables(i,k)|AFSC(i,k) #EQ# 26:
      planned(i,k));
Weather = @sum(variables(i,k)|AFSC(i,k) #EQ# 27:
      planned(i,k));
AllOthers = @sum(variables(i,k)|AFSC(i,k) #EQ# 28:
      planned(i,k));
```

```
! Marisha Kinkle
! Stage 2 Period 3;
model:
title Sample AFSC;
sets:
      student:quality, CommProb, FTProb, sex, race, ethnicity, ScholStat,
tuition;
      choice;
      variables(student,choice): planned, AFSC ;
      jobs: Required;

endsets

data:
      choice = 1, 2, 3, 4;           !2 possible AFSC choices;
      student = @OLE('stage2prac.xls', 'student');
      AFSC = @OLE('stage2prac.xls', 'AFSC');
      quality = @OLE('stage2prac.xls', 'quality');
      sex = @OLE('stage2prac.xls', 'sex');
      race = @OLE('stage2prac.xls', 'race');
      ethnicity = @OLE('stage2prac.xls', 'ethnicity');
      CommProb = @OLE('stage2prac.xls', 'CommProb');
      FTProb = @OLE('stage2prac.xls', 'FTProb');
      ScholStat = @OLE('stage2prac.xls', 'ScholStat');
      tuition = @OLE('stage2prac.xls','tuition');
      Required = 10 0 15 39 13 2 2 3 5 6 11 5 33 2 53 2 30 10 14 8 4 7 0 0 0
0 0 196;
      !Required = 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0;
      TotalRequired = 1261;
      @OLE('stage2results.xls', 'planned3')=planned;
      @OLE('stage2results.xls','student')=student;
      @OLE('stage2results.xls','quality')=quality;
      @OLE('stage2results.xls','ScholStat')=ScholStat;
      @OLE('stage2results.xls','tuition')=tuition;
enddata


max= @sum(student(i):
      @sum(choice(j):
      planned(i,j)*quality(i)));

!Constraints;
@for(jobs(j):                               !Meet AFSC requirements;
@sum(variables(i,k)|AFSC(i,k) #EQ# j:
      planned(i,k))> Required(j));

@sum(variables(i,k)|sex(i) #EQ# 1:          !At least 20% female;
              planned(i,k)) >
               TotalRequired*0.2;

@sum(variables(i,k)|race(i) #eq# 1:         !At least 1% American Indian;
               planned(i,k)) >
               TotalRequired*0.01;

@sum(variables(i,k)|race(i) #eq# 2 #OR# race(i) #eq# 5:     !At least 6%
Asian or Pacific Islander;
               planned(i,k)) >
```

```
                TotalRequired*0.06;
@sum(variables(i,k)|race(i) #eq# 3:       !At least 8% Black;
                planned(i,k)) >
                TotalRequired*0.08;

@sum(variables(i,k)|ethnicity(i) #EQ# 1:  !At least 5% Hispanic;
                planned(i,k)) >
                TotalRequired*0.05;


@for(student(i):                          !No more than one job per student;
      @sum(choice(k):
            planned(i,k)) < 1);
@for(variables(i,k):                      !Don't assign an AFSC that isn't a
choice;
      planned(i,k) < AFSC(i,k));

@sum(variables(i,k):
      planned(i,k)) < TotalRequired;




FinancialMgmt = @sum(variables(i,k)|AFSC(i,k) #EQ# 1:
      planned(i,k));
CostAnalyst = @sum(variables(i,k)|AFSC(i,k) #EQ# 2:
      planned(i,k));
Contracting = @sum(variables(i,k)|AFSC(i,k) #EQ# 3:
      planned(i,k));
CyberSpaceWarfare = @sum(variables(i,k)|AFSC(i,k) #EQ# 4:
      planned(i,k));
AeroEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 5:
      planned(i,k));
CEArch = @sum(variables(i,k)|AFSC(i,k) #EQ# 6:
      planned(i,k));
AstoEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 7:
      planned(i,k));
BehavSciHumanFac = @sum(variables(i,k)|AFSC(i,k) #EQ# 8:
      planned(i,k));
ChemistBiologist = @sum(variables(i,k)|AFSC(i,k) #EQ# 9:
      planned(i,k));
CECivil = @sum(variables(i,k)|AFSC(i,k) #EQ# 10:
      planned(i,k));
CompEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 11:
      planned(i,k));
OSI = @sum(variables(i,k)|AFSC(i,k) #EQ# 12:
      planned(i,k));
ElecEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 13:
      planned(i,k));
CEElecEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 14:
      planned(i,k));
AcqMngmt = @sum(variables(i,k)|AFSC(i,k) #EQ# 15:
      planned(i,k));
CEEnvEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 16:
      planned(i,k));
ProjectEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 17:
      planned(i,k));
```

```
CEGenEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 18:
      planned(i,k));
OpsAnalyst = @sum(variables(i,k)|AFSC(i,k) #EQ# 19:
      planned(i,k));
MechEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 20:
      planned(i,k));
CEMechEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 21:
      planned(i,k));
PhysicsNucEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 22:
      planned(i,k));
Pilot = @sum(variables(i,k)|AFSC(i,k) #EQ# 23:
      planned(i,k));
Navigator = @sum(variables(i,k)|AFSC(i,k) #EQ# 24:
      planned(i,k));
ABM = @sum(variables(i,k)|AFSC(i,k) #EQ# 25:
      planned(i,k));
RemotePilot = @sum(variables(i,k)|AFSC(i,k) #EQ# 26:
      planned(i,k));
Weather = @sum(variables(i,k)|AFSC(i,k) #EQ# 27:
      planned(i,k));
AllOthers = @sum(variables(i,k)|AFSC(i,k) #EQ# 28:
      planned(i,k));

FemaleOfficers = @sum(variables(i,k)|sex(i) #EQ# 1:
                  planned(i,k));

AmericanIndianOfficerCandidates = @sum(variables(i,k)|race(i) #EQ# 1:
                  planned(i,k));
AsianorPacificIslanderCandidates = @sum(variables(i,k)|race(i) #eq# 2 #OR#
race(i) #eq# 5:
               planned(i,k));

BlackOfficerCandidates = @sum(variables(i,k)|race(i) #eq# 3:
               planned(i,k));
MultiracialOfficerCandidates = @sum(variables(i,k)|race(i) #eq# 4:
               planned(i,k));

HispanicOfficersCandidates = @sum(variables(i,k)|ethnicity(i) #EQ# 1:
               planned(i,k));
TotalOfficers = @sum(variables(i,k):
                  planned(i,k));
TuitionCost = @sum(variables(i,k):
      planned(i,k)*tuition(i)*ScholStat(i));
@for(variables(i,j):                              !0-1 variables;
      @bin(planned(i,j)));
```

```
! Marisha Kinkle
! Stage 2 Period 2;
model:
title Sample AFSC;
sets:
      student:quality, CommProb, FTProb, sex, race, ethnicity, ScholStat,
tuition;
      choice;
      variables(student,choice): planned, AFSC ;
      jobs: Required;
```

```
endsets

data:
      choice = 1, 2, 3, 4;          !2 possible AFSC choices;
      student = @OLE('stage2prac.xls', 'student');
      AFSC = @OLE('stage2prac.xls', 'AFSC');
      quality = @OLE('stage2prac.xls', 'quality');
      sex = @OLE('stage2prac.xls', 'sex');
      race = @OLE('stage2prac.xls', 'race');
      ethnicity = @OLE('stage2prac.xls', 'ethnicity');
      CommProb = @OLE('stage2prac.xls', 'CommProb');
      FTProb = @OLE('stage2prac.xls', 'FTProb');
      ScholStat = @OLE('stage2prac.xls', 'ScholStat');
      tuition = @OLE('stage2prac.xls','tuition');
      @OLE('stage2results.xls','tuition')=tuition;
      Required = 10 0 15 39 13 2 2 3 5 6 11 5 33 2 53 2 30 10 14 8 4 7 0 0 0
0 0 0;
      !Required = 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0;
      TotalRequired = 1324;
      @OLE('stage2results.xls', 'planned2')=planned;
      @OLE('stage2results.xls','student')=student;
      @OLE('stage2results.xls','quality')=quality;
      @OLE('stage2results.xls','ScholStat')=ScholStat;
enddata


max= @sum(student(i):
      @sum(choice(j):
      planned(i,j)*quality(i)*CommProb(i)));

!Constraints;
@for(jobs(j):                              !Meet AFSC requirements;
@sum(variables(i,k)|AFSC(i,k) #EQ# j:
      planned(i,k))> Required(j));

@sum(variables(i,k)|sex(i) #EQ# 1:         !At least 20% female;
              planned(i,k)) >
               TotalRequired*0.2;

@sum(variables(i,k)|race(i) #eq# 1:        !At least 1% American Indian;
               planned(i,k)) >
               TotalRequired*0.01;

@sum(variables(i,k)|race(i) #eq# 2 #OR# race(i) #eq# 5:     !At least 6%
Asian or Pacific Islander;
               planned(i,k)) >
               TotalRequired*0.06;
@sum(variables(i,k)|race(i) #eq# 3:        !At least 8% Black;
               planned(i,k)) >
               TotalRequired*0.08;

@sum(variables(i,k)|ethnicity(i) #EQ# 1:   !At least 5% Hispanic;
               planned(i,k)) >
               TotalRequired*0.05;


@for(student(i):                           !No more than one job per student;
```

```
      @sum(choice(k):
            planned(i,k)) < 1);
@for(variables(i,k):                         !Don't assign an AFSC that isn't a
choice;
      planned(i,k) < AFSC(i,k));

@sum(variables(i,k):
      planned(i,k)) < TotalRequired;




FinancialMgmt = @sum(variables(i,k)|AFSC(i,k) #EQ# 1:
      planned(i,k));
CostAnalyst = @sum(variables(i,k)|AFSC(i,k) #EQ# 2:
      planned(i,k));
Contracting = @sum(variables(i,k)|AFSC(i,k) #EQ# 3:
      planned(i,k));
CyberSpaceWarfare = @sum(variables(i,k)|AFSC(i,k) #EQ# 4:
      planned(i,k));
AeroEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 5:
      planned(i,k));
CEArch = @sum(variables(i,k)|AFSC(i,k) #EQ# 6:
      planned(i,k));
AstoEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 7:
```

```
! Marisha Kinkle
! Stage 2 Period 1;
model:
title Sample AFSC;
sets:
      student:quality, CommProb, FTProb, sex, race, ethnicity, ScholStat,
tuition;
      choice;
      variables(student,choice): planned, AFSC ;
      jobs: Required;

endsets

data:
      choice = 1, 2, 3, 4;          !2 possible AFSC choices;
      student = @OLE('stage2prac.xls', 'student');
      AFSC = @OLE('stage2prac.xls', 'AFSC');
      quality = @OLE('stage2prac.xls', 'quality');
      sex = @OLE('stage2prac.xls', 'sex');
      race = @OLE('stage2prac.xls', 'race');
      ethnicity = @OLE('stage2prac.xls', 'ethnicity');
      CommProb = @OLE('stage2prac.xls', 'CommProb');
      FTProb = @OLE('stage2prac.xls', 'FTProb');
      ScholStat = @OLE('stage2prac.xls', 'ScholStat');
      Tuition = @OLE('stage2prac.xls', 'tuition');
      Required = 10 0 15 39 13 2 2 3 5 6 11 5 33 2 53 2 30 10 14 8 4 7 0 0 0
0 0 0;
      !Required = 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0;
      TotalRequired = 1390;
```

```
      @OLE('stage2results.xls', 'planned1')=planned;
      @OLE('stage2results.xls','student')=student;
      @OLE('stage2results.xls','quality')=quality;
      @OLE('stage2results.xls','ScholStat')=ScholStat;
      @OLE('stage2results.xls','tuition')=tuition;

enddata


max= @sum(student(i):
      @sum(choice(j):
      planned(i,j)*quality(i)*FTProb(i)));

!Constraints;
@for(jobs(j):
@sum(variables(i,k)|AFSC(i,k) #EQ# j:
      planned(i,k))> Required(j));                          !Meet AFSC
requirements;




!@sum(variables(i,k)|sex(i) #EQ# 1:       !At least 20% female;
 !              planned(i,k)) >
               TotalRequired*0.2;

!@sum(variables(i,k)|race(i) #eq# 1:            !At least 1% American Indian;
 !               planned(i,k)) >
               TotalRequired*0.01;

!@sum(variables(i,k)|race(i) #eq# 2 #OR# race(i) #eq# 5:    !At least 6%
Asian or Pacific Islander;
 !               planned(i,k)) >
               TotalRequired*0.06;
!@sum(variables(i,k)|race(i) #eq# 3:            !At least 8% Black;
 !               planned(i,k)) >
               TotalRequired*0.08;

!@sum(variables(i,k)|ethnicity(i) #EQ# 1: !At least 5% Hispanic;
 !               planned(i,k)) >
               TotalRequired*0.05;


@for(student(i):                          !No more than one job per student;
      @sum(choice(k):
            planned(i,k)) < 1);
@for(variables(i,k):                      !Don't assign an AFSC that isn't a
choice;
      planned(i,k) < AFSC(i,k));

@sum(variables(i,k):
      planned(i,k)) < TotalRequired;




FinancialMgmt = @sum(variables(i,k)|AFSC(i,k) #EQ# 1:
```

```
      planned(i,k));
CostAnalyst = @sum(variables(i,k)|AFSC(i,k) #EQ# 2:
      planned(i,k));
Contracting = @sum(variables(i,k)|AFSC(i,k) #EQ# 3:
      planned(i,k));
CyberSpaceWarfare = @sum(variables(i,k)|AFSC(i,k) #EQ# 4:
      planned(i,k));
AeroEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 5:
      planned(i,k));
CEArch = @sum(variables(i,k)|AFSC(i,k) #EQ# 6:
      planned(i,k));
AstoEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 7:
      planned(i,k));
BehavSciHumanFac = @sum(variables(i,k)|AFSC(i,k) #EQ# 8:
      planned(i,k));
ChemistBiologist = @sum(variables(i,k)|AFSC(i,k) #EQ# 9:
      planned(i,k));
CECivil = @sum(variables(i,k)|AFSC(i,k) #EQ# 10:
      planned(i,k));
CompEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 11:
      planned(i,k));
OSI = @sum(variables(i,k)|AFSC(i,k) #EQ# 12:
      planned(i,k));
ElecEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 13:
      planned(i,k));
CEElecEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 14:
      planned(i,k));
AcqMngmt = @sum(variables(i,k)|AFSC(i,k) #EQ# 15:
      planned(i,k));
CEEnvEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 16:
      planned(i,k));
ProjectEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 17:
      planned(i,k));
CEGenEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 18:
      planned(i,k));
OpsAnalyst = @sum(variables(i,k)|AFSC(i,k) #EQ# 19:
      planned(i,k));
MechEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 20:
      planned(i,k));
CEMechEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 21:
      planned(i,k));
PhysicsNucEng = @sum(variables(i,k)|AFSC(i,k) #EQ# 22:
      planned(i,k));
Pilot = @sum(variables(i,k)|AFSC(i,k) #EQ# 23:
      planned(i,k));
Navigator = @sum(variables(i,k)|AFSC(i,k) #EQ# 24:
      planned(i,k));
ABM = @sum(variables(i,k)|AFSC(i,k) #EQ# 25:
      planned(i,k));
RemotePilot = @sum(variables(i,k)|AFSC(i,k) #EQ# 26:
      planned(i,k));
Weather = @sum(variables(i,k)|AFSC(i,k) #EQ# 27:
      planned(i,k));
AllOthers = @sum(variables(i,k)|AFSC(i,k) #EQ# 28:
      planned(i,k));

FemaleOfficers = @sum(variables(i,k)|sex(i) #EQ# 1:
```

```
                  planned(i,k));

AmericanIndianOfficerCandidates = @sum(variables(i,k)|race(i) #EQ# 1:
                  planned(i,k));
AsianorPacificIslanderCandidates = @sum(variables(i,k)|race(i) #eq# 2 #OR#
race(i) #eq# 5:
               planned(i,k));

BlackOfficerCandidates = @sum(variables(i,k)|race(i) #eq# 3:
               planned(i,k));
MultiracialOfficerCandidates = @sum(variables(i,k)|race(i) #eq# 4:
               planned(i,k));

HispanicOfficersCandidates = @sum(variables(i,k)|ethnicity(i) #EQ# 1:
               planned(i,k));

TotalOfficers = @sum(variables(i,k):
                  planned(i,k));
TuitionCost = @sum(variables(i,k):
      planned(i,k)*tuition(i)*ScholStat(i));

@for(variables(i,j):                            !0-1 variables;
      @bin(planned(i,j)));
```

Appendix B. Stage Two: Field Training and Commissioned Logistic Regression Results

Table 22. FT Completion Forward Selection Stepwise Logistic Regression Results

Region						0.119	0.127
AS Level					0.053	0.053	0.056
Sex							
Race							
Tech Major				0.015	0.013	0.014	0.013
Cumulative GPA	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Scholarship Status			0.000	0.000	0.000	0.000	0.000
SAT-R							
AFPFT		0.000	0.000	0.000	0.000	0.000	0.000
Mil Experience							0.137
Log-Likelihood	-2291.058	-2249.535	-2229.340	-2226.405	-2224.629	-2219.625	-2218.685
Pearson Test (p-value)	0.000	0.122	0.370	0.647	0.597	0.860	0.867
Deviance Test (p-value)	0.456	1.000	1.000	1.000	1.000	1.000	1.000
Hosmer-Lemeshow (p-value)	0.419	0.639	0.043	0.034	0.026	0.057	0.031
Brown: general alt. (p-value)	0.175	0.000	0.000	0.000	0.000	0.000	0.000
Brown: symmetric alt. (p-value)	0.114	0.002	0.000	0.000	0.000	0.000	0.000

Table 23. FT Completion Backward Elimination Stepwise Logistic Regression Results

Variable	**Model 2.1**	**Model 2.2**	**Model 2.3**	**Model 2.4**	**Model 2.5**
Region	0.150	0.150	0.138	0.127	0.119
AS Level	0.058	0.059	0.059	0.056	0.053
Sex	0.901				
Race	0.722	0.728			
Tech Major	0.025	0.025	0.025	0.013	0.014
Cumulative GPA	0.000	0.000	0.000	0.000	0.000
Scholarship Status	0.000	0.000	0.000	0.000	0.000
SAT-R	0.602	0.608	0.573		
AFPFT	0.000	0.000	0.000	0.000	0.000
Mil Experience	0.135	0.137	0.135	0.137	
Log-Likelihood	-2214.213	-2214.215	-2218.436	-2218.685	-2219.625
Pearson Test (p-value)	0.868	0.870	0.917	0.867	0.860
Deviance Test (p-value)	1	1.000	1.000	1.000	1.000
Hosmer-Lemeshow (p-value)	0.123	0.084	0.031	0.031	0.057
Brown: general alt. (p-value)	0.000	0.000	0.000	0.000	0.000
Brown: symmetric alt. (p-value)	0.001	0.001	0.000	0.000	0.000

Table 24. FT Completion Mixed Stepwise Logistic Regression Results

Variable	Model 3.1	Model 3.2	Model 3.3	Model 3.4	Model 3.5	Model 3.6	Model 3.7
Region						0.119	0.127
AS Level					0.053	0.053	0.056
Sex							
Race							
Tech Major				0.015	0.013	0.014	0.013
Cumulative GPA	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Scholarship Status			0.000	0.000	0.000	0.000	0.000
SAT-R							
AFPFT		0.000	0.000	0.000	0.000	0.000	0.000
Mil Experience							0.137
Log-Likelihood	-2291.058	-2249.535	-2229.340	-2226.405	-2224.629	-2219.625	-2218.685
Pearson Test (p-value)	0.000	0.122	0.370	0.647	0.597	0.860	0.867
Deviance Test (p-value)	0.456	1.000	1.000	1.000	1.000	1.000	1.000
Hosmer-Lemeshow (p-value)	0.419	0.639	0.043	0.034	0.026	0.057	0.031
Brown: general alt. (p-value)	0.175	0.000	0.000	0.000	0.000	0.000	0.000
Brown: symmetric alt. (p-value)	0.114	0.002	0.000	0.000	0.000	0.000	0.000

Table 25. Commissioned Forward Selection Stepwise Logistic Regression

Variable	Model 10.1	Model 10.2
Region	0.000	0.000
Ethnicity		0.023
Mil Experience	0.007	0.007
*Region*Ethnicity*		
*Region*Mil Experience*	0.160	0.148
*Ethnicity*MilExperience*		
Log-Likelihood	-2922.031	-2919.185
Pearson Test (p-value)	0.242	0.410
Deviance Test (p-value)	0.253	0.238
Hosmer-Lemeshow (p-value)	0.926	0.963
Brown: general alt. (p-value)	0.242	0.929
Brown: symmetric alt. (p-value)	0.095	0.822

Table 26. Commissioned Backward Elimination Stepwise Logistic Regression

Variable	Model 11.1	Model 11.2	Model 11.3	Model 11.4
Region	0.000	0.000	0.000	0.000
Ethnicity	0.011	0.054	0.023	0.024
Mil Experience	0.014	0.014	0.007	0.004
*Region*Ethnicity*	0.635			
*Region*Mil Experience*	0.129	0.126		
*Ethnicity*MilExperience*	0.289	0.276	0.148	
Log-Likelihood	-2918.142	-2918.268	-2918.388	-2920.178
Pearson Test (p-value)	0.269	0.337	0.288	0.325
Deviance Test (p-value)	0.260	0.329	0.250	0.225
Hosmer-Lemeshow (p-value)	0.955	0.978	0.821	0.804
Brown: general alt. (p-value)	0.274	0.642	0.210	0.530
Brown: symmetric alt. (p-value)	0.427	0.706	0.521	0.451

Table 27. Commissioned Mixed Stepwise Logistic Regression

Variable	Model 12.1
Region	0.000
Ethnicity	0.023
Mil Experience	0.007
*Region*Ethnicity*	
*Region*Mil Experience*	0.148
*Ethnicity*MilExperience*	
Log-Likelihood	-2919.185
Pearson Test (p-value)	0.410
Deviance Test (p-value)	0.276
Hosmer-Lemeshow (p-value)	0.963
Brown: general alt. (p-value)	0.929
Brown: symmetric alt. (p-value)	0.822

Nominal Logistic Fit for FT Comp

Converged in Gradient, 6 iterations

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	147.4597	6	294.9194	<.0001*
Full	2218.2135			
Reduced	2365.6732			

RSquare (U)	0.0623
AICc	4450.44
BIC	4498.59
Observations (or Sum Wgts)	7189

Measure	Training	Definition
Entropy RSquare	0.0623	1-Loglike(model)/Loglike(0)
Generalized R-Square	0.0834	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.3086	∑ -Log(ρ[j])/n
RMSE	0.2960	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.1750	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.1021	∑ (ρ[j]≠ρMax)/n
N	7189	n

Lack Of Fit

Source	DF	-LogLikelihood	ChiSquare
Lack Of Fit	6380	2100.9506	4201.901
Saturated	6386	117.2630	**Prob>ChiSq**
Fitted	6	2218.2135	1.0000

Parameter Estimates

Term	Estimate	Std Error	ChiSquare
Intercept	5.89016682	0.6631271	78.90
Tech major	0.20009948	0.0814118	6.04
CGPA	-0.7896811	0.0814266	94.05
Schol Stat	-0.6756668	0.0926124	53.23
AFPFT Score	-0.0589259	0.0071116	68.66
(CGPA-3.10001)*(Schol Stat-0.7581)	-0.4424458	0.1645472	7.23
(AFPFT Score-90.4031)*(AFPFT Score-90.4031)	-0.0008343	0.0003002	7.72

For log odds of 0/1

Effect Likelihood Ratio Tests

Source	Nparm	DF	L-R ChiSquare	Prob>ChiSq
Tech major	1	1	6.0317592	0.0141*
CGPA	1	1	93.4039986	<.0001*
Schol Stat	1	1	49.8978609	<.0001*
AFPFT Score	1	1	72.7000988	<.0001*
CGPA*Schol Stat	1	1	7.1792443	0.0074*
AFPFT Score*AFPFT Score	1	1	8.73004611	0.0031*

Nominal Logistic Fit for Comm

Converged in Gradient, 5 iterations

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	37.0387	4	74.07737	<.0001*
Full	2927.1301			
Reduced	2964.1688			

RSquare (U)	0.0125
AICc	5864.27
BIC	5898.05
Observations (or Sum Wgts)	6357

Measure	Training	Definition
Entropy RSquare	0.0125	1-Loglike(model)/Loglike(0)
Generalized R-Square	0.0191	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.4605	∑ -Log(ρ[j])/n
RMSE	0.3790	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.2875	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.1767	∑ (ρ[j]≠ρMax)/n
N	6357	n

Lack Of Fit

Source	DF	-LogLikelihood	ChiSquare
Lack Of Fit	2113	1082.5938	2165.188
Saturated	2117	1844.5363	**Prob>ChiSq**
Fitted	4	2927.1301	0.2100

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq	Lower 95%
Intercept	-1.834966	0.2326183	62.23	<.0001*	-2.2929827
Region	0.22514274	0.0299293	56.59	<.0001*	0.1666873
Military Experience	-0.4319335	0.1488619	8.42	0.0037*	-0.7336016
Ethnicity	-0.3640257	0.1577673	5.32	0.0210*	-0.6843752
Cum GPA	-0.0854118	0.0702305	1.48	0.2239	-0.2229622

For log odds of 0/1

Effect Wald Tests

Source	Nparm	DF	Wald ChiSquare	Prob>ChiSq
Region	1	1	56.5877123	<.0001*
Military Experience	1	1	8.41912303	0.0037*
Ethnicity	1	1	5.32390451	0.0210*
Cum GPA	1	1	1.47905234	0.2239

Effect Likelihood Ratio Tests

Source	Nparm	DF	L-R ChiSquare	Prob>ChiSq
Region	1	1	57.9042747	<.0001*
Military Experience	1	1	9.23746248	0.0024*
Ethnicity	1	1	5.75664728	0.0164*
Cum GPA	1	1	1.47722798	0.2242

Odds Ratios
For Comm odds of 0 versus 1

Unit Odds Ratios
Per unit change in regressor

Term	Odds Ratio	Lower 95%	Upper 95%	Reciprocal
Region	1.252501	1.181385	1.328474	0.7984022
Military Experience	0.649253	0.480176	0.861593	1.5402326
Ethnicity	0.694873	0.504405	0.937421	1.4391112
Cum GPA	0.918134	0.800145	1.05378	1.0891655

Range Odds Ratios
Per change in regressor over entire range

Term	Odds Ratio	Lower 95%	Upper 95%	Reciprocal
Region	1.964874	1.648823	2.344547	0.5089384
Military Experience	0.649253	0.480176	0.861593	1.5402326
Ethnicity	0.694873	0.504405	0.937421	1.4391112
Cum GPA	0.757609	0.484505	1.185596	1.3199426

Appendix C. Stage One Probabilities

Table 28. Probabilities Using FY03 Data by Individual Composite Score

Using FY03 Applicant Data										
Lower	**Upper**	**Tot Eligible**	**# Offer**	**# Accept**	**# Comm (Acc)**	**# Comm (Offer Not Acc)**	**Total Comm**	**P(Offer — Apply)**	**P(Acceptance— Offered)**	**P(Graduation—Accept Scholarship)**
0	39	1	0	0	0		0	0	0	0
40	44	0	0	0	0		0	0	0%	0%
45	49	1	1	1	0		0	1	100%	0%
50	54	21	2	2	0		0	0.095238	100%	0%
55	59	94	18	15	7		7	0.191489	83%	47%
60	64	414	57	49	19		19	0.137681	86%	39%
65	69	1169	283	222	81		81	0.242087	78%	36%
70	74	1493	484	354	146		146	0.32418	73%	41%
75	79	1278	507	341	149		149	0.396714	67%	44%
80	84	963	427	274	101		101	0.443406	64%	37%
85	89	538	287	183	76		76	0.533457	64%	42%
90	94	200	147	77	38		38	0.735	52%	49%
95	100	25	16	5	2		2	0.64	31%	40%
	Total	**6197**	**2229**	**1523**	**619**	**0**	**619**			

Table 29. Probabilities Using FY04 Data by Individual Composite Score

Using FY04 Applicant Data										
Lower	**Upper**	**Tot Eligible**	**# Offer**	**# Accept**	**# Comm (Acc)**	**# Comm (Offer Not Acc)**	**Total Comm**	**P(Offer — Apply)**	**P(Acceptance— Offered)**	**P(Graduation—Accept Scholarship)**
0	39	3	0	0	0	0	0	0%	0%	0%
40	44	4	0	0	0	0	0	0%	0%	0%
45	49	1	0	0	0	0	0	0%	0%	0%
50	54	15	0	0	0	0	0	0%	0%	0%
55	59	36	1	0	0	0	0	3%	0%	0%
60	64	82	6	3	0	1	1	7%	50%	0%
65	69	223	38	11	7	7	14	17%	29%	64%
70	74	303	50	22	5	6	11	17%	44%	23%
75	79	344	66	20	10	16	26	19%	30%	50%
80	84	274	52	15	5	15	20	19%	29%	33%
85	89	171	41	16	3	6	9	24%	39%	19%
90	94	86	28	7	2	3	5	33%	25%	29%
95	100	14	3	0	0	1	1	21%	0%	0%
	Total	**1556**	**285**	**94**	**32**	**55**	**87**			

Table 30. Historical Average Probabilities Using FY01 - FY06 Data by Individual Composite Score

	Probability of being Offered a Scholarship by Individual Composite Score													
App Year	***Unemp Rate***	***0-39***	***40-44***	***45-49***	***50-54***	***55-59***	***60-64***	***65-69***	***70-74***	***75-79***	***80-84***	***85-89***	***90-94***	***95-100***
2001	4.7	0.98	0.25	0.80	0.67	0.52	0.62	0.75	0.83	0.88	0.90	0.93	0.95	0.90
2002	5.8	1.00	0.00	0.40	0.50	0.36	0.43	0.53	0.60	0.69	0.71	0.77	0.80	0.77
2003	6	0.00	0.00	1.00	0.15	0.27	0.21	0.37	0.50	0.59	0.64	0.72	0.90	0.89
2004	5.5	0.00	0.00	1.00	0.10	0.19	0.14	0.24	0.32	0.40	0.44	0.53	0.74	0.64
2005	5.1													
2006	4.6	0.38	0.21	0.55	0.38	0.28	0.28	0.41	0.55	0.59	0.62	0.74	0.86	0.89
2007	4.6	0.00	0.67	0.67	0.39	0.30	0.33	0.51	0.63	0.65	0.77	0.79	0.93	1.00
2008	5.8	0.00	0.00	0.00	0.35	0.37	0.41	0.54	0.66	0.75	0.81	0.87	0.93	0.92
2009	9.3	0.00	0.00	1.00	0.28	0.16	0.25	0.40	0.55	0.65	0.70	0.77	0.88	0.94
2010	9.6	0.00	0.00	0.00	0.00	0.06	0.13	0.26	0.38	0.44	0.54	0.63	0.74	0.76
Average:		0.26	0.13	0.60	0.31	0.28	0.31	0.45	0.56	0.63	0.68	0.75	0.86	0.86

	Probability of being Graduating (w/in 4 years) given Acceptance of Scholarship by Individual Composite Score													
App Year	***Unemp Rate***	***0-39***	***40-44***	***45-49***	***50-54***	***55-59***	***60-64***	***65-69***	***70-74***	***75-79***	***80-84***	***85-89***	***90-94***	***95-100***
2001	4.7	0.44	0.00	0.75	0.23	0.24	0.36	0.37	0.41	0.45	0.42	0.43	0.53	0.57
2002	5.8	0.40	0.00	0.00	0.40	0.34	0.38	0.35	0.41	0.43	0.43	0.45	0.37	0.60
2003	6	0.00	0.00	0.00	0.00	0.39	0.35	0.30	0.30	0.30	0.24	0.27	0.26	0.13
2004	5.5	0.00	0.00	0.00	0.00	0.47	0.39	0.36	0.41	0.44	0.37	0.42	0.49	0.40
2005	5.1													
2006	4.6	0.00	0.20	0.37	0.36	0.30	0.29	0.28	0.26	0.38	0.39	0.39	0.45	0.25
Average:		0.17	0.04	0.22	0.20	0.35	0.35	0.33	0.36	0.40	0.37	0.39	0.42	0.39

Table 31. Probabilities Using FY03 Data by SAT Score

Using FY03 Applicant Data										
Lower	**Upper**	**Tot Eligible**	**# Offer**	**# Accept**	**# Comm (Acc)**	**# Comm (Offer Not Acc)**	**Total Comm**	**P(Offer— Apply)**	**P(Acceptance— Offered)**	**P(Graduation—Accept Scholarship)**
0	999	0	0	0	0	0	0		0	0
1000	1049	0	0	0	0	0	0		0%	0%
1050	1099	1	0	0	0	0	0	0%	0%	0%
1100	1149	1450	381	303	129	1	130	26%	80%	43%
1150	1199	1041	312	230	94	1	95	30%	74%	41%
1200	1249	1083	347	231	95	2	97	32%	67%	41%
1250	1299	910	346	208	97	3	100	38%	60%	47%
1300	1349	841	375	213	91	1	92	45%	57%	43%
1350	1399	361	187	98	45	0	45	52%	52%	46%
1400	1449	263	135	80	32	1	33	51%	59%	40%
1450	1499	154	91	49	27	0	27	59%	54%	55%
1500	1549	59	33	13	4	0	4	56%	39%	31%
1550	1600	34	22	8	3	0	3	65%	36%	38%
	Total	**6197**	**2229**	**1433**	**617**	**9**	**626**			

Table 32. Probabilities Using FY04 Data by SAT Score

Using FY04 Applicant Data										
Lower	**Upper**	**Tot Eligible**	**# Offer**	**# Accept**	**# Comm (Acc)**	**# Comm (Offer Not Acc)**	**Total Comm**	**P(Offer— Apply)**	**P(Acceptance— Offered)**	**P(Graduation—Accept Scholarship)**
0	999	0	0	0	0	0	0		0	0
1000	1049	0	0	0	0	0	0		0%	0%
1050	1099	37	0	0	0	0	0	0%	0%	0%
1100	1149	307	45	18	7	6	13	15%	40%	39%
1150	1199	250	38	12	6	8	14	15%	32%	50%
1200	1249	228	49	20	7	11	18	21%	41%	35%
1250	1299	243	48	12	4	12	16	20%	25%	33%
1300	1349	241	53	14	4	8	12	22%	26%	29%
1350	1399	104	25	10	3	3	6	24%	40%	30%
1400	1449	79	19	5	0	1	1	24%	26%	0%
1450	1499	41	8	0	0	4	4	20%	0%	0%
1500	1549	16	5	2	1	0	1	31%	40%	50%
1550	1600	10	3	1	0	0	0	30%	33%	0%
	Total	**1556**	**293**	**94**	**32**	**53**	**85**			

Table 33. Historical Average Probabilities Using FY01 - FY06 Data by SAT Score

Using FY01-06 Applicant Data										
Lower	**Upper**	**Tot Eligible**	**# Offer**	**# Accept**	**# Comm (Acc)**	**# Comm (Offer Not Acc)**	**Total Comm**	**P(Offer— Apply)**	**P(Acceptance— Offered)**	**P(Graduation—Accept Scholarship)**
0	999	0	0	0	0	0	0		0	0
1000	1049	0	0	0	0	0	0		0%	0%
1050	1099	1259	581	415	151	76	227	46%	71%	36%
1100	1149	5280	1929	1796	690	332	1022	37%	93%	38%
1150	1199	3746	1479	1262	534	216	750	39%	85%	42%
1200	1249	3644	1628	1237	488	181	669	45%	76%	39%
1250	1299	3108	1500	1053	445	185	630	48%	70%	42%
1300	1349	2949	1521	997	435	143	578	52%	66%	44%
1350	1399	1397	792	487	230	49	279	57%	61%	47%
1400	1449	890	526	323	135	23	158	59%	61%	42%
1450	1499	476	279	137	72	22	94	59%	49%	53%
1500	1549	209	134	59	23	8	31	64%	44%	39%
1550	1600	101	66	32	16	2	18	65%	48%	50%
	Total	**23059**	**10435**	**7798**	**3219**	**1237**	**4456**			

Table 34. Stage One Tuition Costs by Scholarship Type

Type	Avg Cost	Offer Rate	Acceptance Rate
1	$9,391	10.2%	40.0%
2	$5,305	36.5%	45.8%
7	$3,969	53.2%	21.2%

Appendix D. Thesis Storyboard

A Multi-Stage Model for Air Force Reserve Officer Training Corps Officer Candidate Selection

Capt Marisha Kinkle

Committee:

Advisor: Maj Matthew J. Robbins, Ph.D.

Member: Dr. Darryl K. Ahner

Department of Operational Sciences (ENS)

Air Force Institute of Technology

Introduction

The Air Force Reserve Officer Training Corps (AFROTC) faces a declining budget and increased enrollment, creating the necessity for improving officer candidate selection thorough the various stages of its commissioning program. Three critical stages have a major impact on the type of officer AFROTC commission. This research proposes a multi-stage model to evaluate three stages: 1) the high school scholarship allocation process, 2) the in-college scholarship allocation process, and 3) commissioning. Each stage is examined individually so that collectively AFROTC decision makers are able to meet commissioning goals.

Methodology

Stage one involves allocating scholarships to high school candidates using the index policy heuristic. Stage two involves examining which candidates should be awarded an enrollment allocation while taking into account the probabilities of the candidate completing field training (FT) and going on to commission. A logistic regression is used to estimate the probabilities of FT completion and commissioning given a candidate's demographic information and college performance. Stage two is examined using dynamic programming with a knapsack formulation. Stage three involves selecting the most qualified cadets to commission into the USAF and is examined using a knapsack approach.

AIR FORCE INSTITUTE OF TECHNOLOGY

Results and Analysis

Stage One Results

Individual Composite Score										
Data Used	# Apply	# Offers	#Accept	Scholarship Cost	Offer Avg Quality	Percntage Change	# Comm	Comm Avg Quality	Percentage Change	Total Quality
Actual	1620	293	94	$503,626	79		85	79		6,672.50
Historical (2001-2006 Averages)	1620	301	97	$517,342	88	10.1%	85	87	10.8%	7,395.00
Overlap		82			89		23	87		2,001.00
FY04 Data	1620	340	109	$584,374	76	-6.6%	85	74	-6.0%	6,271.30
Overlap		72			79		21	76		1,585.50
FY03 Data	1620	301	97	$517,342	88	7.6%	85	85	8.3%	7,225.00
Overlap		87			87		28	84		2,352.00

SAT										
Data Used	# Apply	# Offers	#Accept	Scholarship Cost	Offer Avg Quality	Percntage Change	# Comm	Comm Avg Quality	Percntage Change	Total Quality
Actual	1552	293	94	$503,626	1266		85	1246		105,910.00
Historical (2001-2006 Averages)	1552	292	94	$503,626	1405	11.0%	85	1373	10.2%	116,705.00
Overlap		85			1279		28	1251		35,028.00
FY04 Data	1552	360	115	$618,749	1275	0.7%	85	1253	0.6%	106,505.00
Overlap		85			1279		28	1251		35,028.00
FY03 Data	1552	292	94	$503,626	1404	10.9%	85	1373	10.2%	116,705.00
Overlap		69			1408		21	1377		28,917.00

Stage Two Results

Basic	Period 3	Percentages	Period 2	Percentages	Period 1	Percentages
Overall Quality Score	99,155		83,790		91,487	
Mean Quality Score	79		77		76	
Diversity						
Female	280	22.2%	294	22.2%	316	22.7%
Hispanic	43	3.4%	45	3.4%	51	3.7%
African American	34	2.7%	36	2.7%	45	3.2%
Asian or Pacific Islander	66	5.2%	65	4.9%	71	5.1%
American Indian	8	0.6%	8	0.6%	8	0.6%
MultiRacial	12	1.0%	12	0.9%	13	0.9%
Projected Tuition Cost	$5,167,053.00		$5,469,853.00		$5,581,150.00	

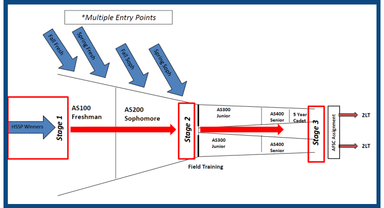
Expanded	Period 3	Percentages	Period 2	Percentages	Period 1	Percentages
Overall Quality Score	97,462	-1.7%	82,522	-1.5%	90,100	-1.5%
Mean Quality Score	77		76		74	
Diversity						
Female	282	22.4%	299	22.6%	315	22.7%
Hispanic	64	5.1%	67	5.1%	70	5.0%
African American	101	8.0%	106	8.0%	112	8.1%
Asian or Pacific Islander	76	6.0%	80	6.0%	84	6.0%
American Indian	13	1.0%	14	1.1%	14	1.0%
MultiRacial	11	0.9%	11	0.8%	11	0.8%
Projected Tuition Cost	$4,951,385.00		$5,186,461.00		$5,500,031.00	

Stage Three Results

	Basic	Extended
Overall Quality Score	109,640.30	109,279.40
Female Officers	14.39%	20.03%
Minority Officers	14.48%	20.03%
Hispanic Officers	4.71%	5.05%

AFSC Allocation				
Career	AFSC	Required	Basic	Extended
Financial Mgmt	65F/W	10	10	10
Cost Analyst	65WX	0	0	0
Contracting	64P	15	15	15
Cyber Space Warfare	17D	39	39	39
Aeronautical Engr	62EXA	13	13	13
C-E (Arch)	32EXA	2	2	2
Astronautical Engr	62EXB	2	2	2
Behav Sci/Human Factors	61BX	3	3	3
Chemist/Biologist	61CX	5	5	5
C-E (Civil)	32EXC	6	6	6
Computer Engr	62EXC	11	11	11
OSI	71Sx	5	5	5
Electrical Engr	62EXE	33	33	33
C-E (EE)	32EXE	2	2	2
Acquisition Mgmt	63A	53	53	53
C-E(Envir)	32EXJ	2	2	2
Project Engr	62EXG	30	30	30
C-E (Gen)	32EXG	10	10	10
Operations Research Analyst	61AX	14	14	14
Mechanical Engr	*62EXH*	8	8	8
C-E (Mech)	32EXF	4	4	4
Physics/Nuclear Engineer	61DX	7	7	7
Pilot	11	511	511	511
Navigator	12	120	120	120
Air Battle Managmt	13	75	75	75
Remote Piloted Aircraft	18	12	12	12
Weather	15W	1	1	1
All Others		195	195	195

Model Framework



Motivation

AFROTC faces a declining budget and increased enrolment, creating the necessity for improving officer candidate selection

Impacts/Contributions

- Stage one offers effective optimization for allocation of scholarships
- Stage two logistic regression analysis allows insights into contributing factors toward FT completion and commissioning
- Stages two and three allow decision makers an effective optimization tool for enrollment allocations and AFSC selection respectively
  - Allows the decision maker to consider other constraints such as diversity or cost
  - Allows for sensitivity analysis of requirements

Collaboration

Headquarters AF – A1

Jeanne M. Holm Center for Officer Accessions & Citizen Development

Air Force Reserve Officer Training Corps

Bibliography

[1] 36-2011, AFROTCI. "Air Force Reserve Officer Training Corps Instruction 36-2011: Cadet Operations", 2010.

[2] 36-2019, AFROTCI. "Air Force Reserve Officer Training Corps Instruction 36-2019: AFROTC Scholarship Programs", 2004.

[3] 36-2101, AFI. "Air Force Instruction 36-2011: Classifying Military Personnel (Officer and Enlisted)", 2010.

[4] Ali, A. I., Jeffery L. Kennington, and Timothy T. Liang. "Assignment with En route training of navy personnel", *Naval research logistics*, 40(5):581–581, 1993. Doi: pmid:.

[5] Bertsekas, Dimitri. *Dynamic Programming & Optimal Control*. Athena Scientific, City, 2005. ISBN 1886529264.

[6] Bertsekas, Dimitri P. and David A. Castanon. "Rollout Algorithms for Stochastic Scheduling Problems", *Journal of Heuristics*, 5(1):89–108, 1999.

[7] Brown, Charles C. "On a goodness of fit test for the logistic model based on score statistics", *Communications in Statistics - Theory and Methods*, 11(10):1087–1105, 1982.

[8] Camarena-Alvarado, Lilia. "Scholarship Allocation: an Optimization Model for Enrollment Management", 2010.

[9] Center, Air Force Personnel. "Interactive Demographic Analysis System (IDEAS)", 2011. URL `http://w11.afpc.randolph.af.mil/vbin/broker8.exe?_program=ideas.IDEAS_Default.sas&_service=prod2pool3&_debug=0`.

[10] Chen, Der-San, Robert G. Batson, and Yu Dang. *Applied Integer Programming: Modeling and Solution*. Wiley, 2010. ISBN 0470373067.

[11] Corps, Air Force Reserve Officer Training. "U.S. Air Force ROTC", 2011. URL `http://www.afrotc.com/learn-about/mission-and-values/`.

[12] Goldman, Charles. *Allocating Scholarship for Army ROTC*. RAND, Santa Monica, CA, 1999.

[13] Guendel, Fredrick. "HQ AFROTC Sponsored Thesis Topic", 2011.

[14] Hallett, David C. "Goodness of Fit Tests in Logistic Regression", 1999.

[15] Hardison, Chaitra M., Carra S. Sims, and Eunice C. Wong. *The Air Force Officer Qualifying Test: Validity, Fairness and Bias*. Rand Publishing, 2010. ISBN 0833047795.

[16] Howard, Ronald. "Dynamic Programming", *Management Science*, 12(5):317–348, 1966.

[17] Hoyer, Mary. "HSSP Scholarship Process", 2011.

[18] Johnson, Joseph G. and Jerome R. Busemeyer. "Multiple-Stage Decision-Making: The Effect of Planning Horizon Length on Dynamic Consistency", 51(2):217–246, 2001.

[19] Kellerer, Hans, Ulrich Pferschy, and David Pisinger. *Knapsack Problems*. Springer, 2004. ISBN 3540402861.

[20] Montgomery, Douglas C., Elizabeth A. Peck, and G. Geoffrey Vining. *Introduction to Linear Regression Analysis (Wiley Series in Probability and Statistics)*. Wiley-Interscience, 2006. ISBN 0471754951.

[21] Muniz, Ingrid. "Enrollment Allocations", 2012.

[22] Raymond, Jonathan D. *Determining the Number of Reenlistments Necessary to Satisfy Future Force Requirements*. Naval Postgraduate School, California, 2006.

[23] Shmueli, Galit, Nitin R. Patel, and Peter C. Bruce. *Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner*. Wiley, 2010. ISBN 0470526823.

Vita

Captain Marisha T. Kinkle completed high school at Wichita Southeast High School in Wichita, KS in 2002. She accomplished her undergraduate studies at Spelman College in 2006 with a Bachelor of Arts degree in Mathematics. Marisha was commissioned into the US Air Force as a crosstown through AFROTC Detachment 165 located at Georgia Institute of Technology.

Captain Kinkle's first assignment was to Air Force Officer Accession and Training Schools, now named the Jeanne M. Holm Center for Officer Accessions and Citizen Development, at Maxwell AFB, Alabama. Initially, Marisha worked at Headquarters Air Force Reserve Officer Training Corps as chief of plans and programs. She planned, coordinated, and directed daily operations for the command section. In 2007, Marisha moved to the Holm Center Commander's Action Group where she developed plans and conducted analysis for Air Force Reserve Officer Training Corps, Air Force Junior Reserve Officer Training Corps, and Officer Training Schools.

In June 2009, Marisha was assigned to the LeMay Center for Doctrine Development and Education which is also located at Maxwell AFB, Alabama. She provided analysis to facilitate educational wargames constructed to support professional military education at Air University.

In August 2010, Marisha entered the Air Force Institute of Technology's Graduate School of Engineering and Management at Wright-Patterson AFB, Ohio. At AFIT, she focused her studies on Decision Analysis in the field of Operations Research. Upon graduation, she will be assigned to the Air Force Material Command's Analysis Branch located at Wright-Patterson AFB, OH.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704–0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202–4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
22–03–2011	Master's Thesis	Aug 2010 – Mar 2012

4. TITLE AND SUBTITLE

A Multi-Stage Optimization Model
for Air Force Reserve Officer Training Corps
Officer Candidate Selection

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Marisha T. Kinkle, Capt, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-OR-MS-ENS-12-16

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Col Frederick I. Guendel
AFROTC/RR
551 E. Maxwell Blvd
Maxwell AFB, AL 36112
(334) 953-5588, frederick.guendel@maxwell.af.mil

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

13. SUPPLEMENTARY NOTES

14. ABSTRACT The Air Force Reserve Officer Training Corps (AFROTC) faces a declining budget and increased enrollment, creating the necessity for improving officer candidate selection thorough the various stages of its commissioning program. Among all stages of the program, the three critical stages affecting the type of officer commissioned are: high school scholarship allocation, enrollment allocation, and AFSC assignment.
This research proposes a multi-stage problem addressing each of the three critical stages. Each stage is examined individually so collectively AFROTC decision makers are able to meet commissioning goals. Stage one involves allocating scholarships to high school candidates using the quiz policy heuristic. Stage two involves examining which candidates should be awarded an enrollment allocation while considering the probabilities of the candidate completing field training (FT) and going on to commission. A logistic regression is used to estimate the probabilities of FT completion and commissioning. Stage two is examined using dynamic programming with a knapsack formulation. Stage three involves selecting the most qualified cadets to commission into the USAF and is examined using a knapsack approach.

15. SUBJECT TERMS

AFROTC, Dynamic Programming, Knapsack, Logistic Regression, Quiz Problem

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Matthew J. Robbins
					19b. TELEPHONE NUMBER *(include area code)*
U	U	U	UU	102	(937) 255–3636 ext 4539; matthew.robbins@afit.edu